Engineering Single-Chain Rop Variants and Combinatorial Libraries

Thesis

Presented in Partial Fulfillment of the Requirements for Graduation *With Distinction* in Chemistry in the Undergraduate College of Arts and Sciences at Ohio State

By

Danielle Marie Williams

Undergraduate Program in Chemistry

The Ohio State University
2012

Thesis Committee:

Dr. Thomas Magliery, Advisor

Dr. Jane Jackman

Dr. Christopher Jaroniec

Abstract

Understanding the basis of protein stability is one of the most profound current fundamental problems in biophysics. A single mutation can cause misfolding or prevent folding, leading to diseases such as cancer and cystic fibrosis. To explore the relationship between protein sequence and function, model proteins can be altered through rational design and characterized to study the effects of the changes. Rop is a small, homodimeric, four-helix bundle and is a well-characterized model protein due to its small size and simplicity. My project involves the engineering of a single-chain Rop that has similar activity and stability with respect to the wild type. Two variants of a single chain mutant with known X-ray crystal structure (PDB ID: 1yo7) were created by making point mutations to the loop sequences, and they differ by the deletion of an isoleucine in one loop. We characterized these proteins using CD, NMR, and crystallography, and we tested them for activity in a GFP-based screen. Both variants are alpha-helical and well-folded, and thermal and urea denaturations have indicated cooperative unfolding. Both variants are significantly more active than 1yo7, comparable to wild type. NMR data also suggest that the proteins are well-folded in solution. Analysis by gel filtration chromatography shows that the proteins exist as monomers in solution, and both variants were crystallized with PEG precipitant. Crystals of the variant without isoleucine diffracted well, and we are solving the structure by molecular replacement. We have also characterized twelve

variants from a combinatorial library of Rop in which five solvent-exposed residues were randomized. This research provides insight into how sequence alters protein stability. Such data will eventually enable us to rationally design proteins as drugs and for other potentially useful applications in industry.

Acknowledgements

I would like to thank my advisor, Dr. Tom Magliery for giving me the opportunity to work in his lab as an undergraduate. I would also like to thank graduate student Shiladitya Sen for training me/putting up with me, and for offering continued guidance and support throughout my time in the lab. Chau Nguyen also helped me get started with these projects, and she deserves a lot of credit for her work on the surface library. Kim Stephany has helped tremendously with the characterization of the surface variants, and Sarah Johnston has been a great friend and coworker on each of the Rop projects. Chunhua Yuan was kind enough to help us take the HSQC scans that are shown in this document, and Dr. Jeremy Beck from Dr. Hadad's lab provided some MD simulations on single-chain variants of interest. I would also like to thank my committee members Dr. Christopher Jaroniec and Dr. Jane Jackman, as well as all members of the Magliery Lab.

I would also like to thank Ohio State's Department of Chemistry and Biochemistry for funding, as well as the College of Arts and Sciences Undergraduate Research Scholarship and Dean's Undergraduate Research Fund. Additionally, the National Institutes of Health funded a large part of this project with the RO1 G838114 grant.

Vita

June 2008……..…….Boardman High School

June 2012……...…..Candidate for B.A. in Chemistry, The Ohio State University

Fields of Study

Major Field: Chemistry

Minor Field: Italian

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Inverse folding problem

Understanding the basis of protein stability is one of the most profound and fundamental problems in biophysics. Christian Anfinson set the stage for what would be countless sequence-stability experiments after his work with ribonuclease in the 1950s. He correlated primary sequence with tertiary protein structure, and proved that all the information necessary to determine the conformation of a protein is encoded in its amino-acid sequence.[1] Significant progress has been made in the field of protein science since Anfinsen's time, but there is still no rigorous method to predict protein structure based on sequence alone. While we have a good knowledge base regarding the forces that govern protein folding and stability, we have not reduced it to practice, and the vast biomedical applications make this field a hot topic of interest.[2,3] Such a complicated prediction becomes even more difficult for mutations or insertions/deletions in proteins.

The enormity of possible protein sequences coupled with the time-consuming methods of biophysical characterization cause considerable difficulty in drawing conclusions about the relationship between protein sequence and structure. Levinthal's paradox states that proteins can, in theory, adopt an infinite number of conformations, but could never visit all of them in the time necessary to fold.[4]

Scientists have pegged proteins as folding through a "funnel," ultimately leading to their lowest energy conformation in a matter of seconds or less.[5] There is currently no model to determine which residues have the greatest effect on protein folding or what certain positions contribute the most to stability. Most experimental and computational methods rely on understanding the structure from the primary sequence. Another more pragmatic approach in modern drug design involves the "inverse folding problem." Given a structure or desired function, is it possible to come up with an amino acid sequence that will yield a protein of that specific fold or similar functionality?[6] Creating and analyzing the nearly infinite amount of possible sequences is impossible, and other techniques must be applied to more rapidly and quantitatively collect data.

1.2 Combinatorial methods and rational design

There are primarily two ways to approach this issue: experimentally or computationally. Recent developments in the latter have generated some success in predicting protein structure. Using an algorithm founded on physical chemical potential functions and stereochemical limits, Mayo and Dahiyat were able to synthesize a well-ordered protein (FSD-1) that matched their target ββα protein motif. The sequence had little similarity with any protein found in nature, but the structure was shown to be in agreement with the desired computational design by nuclear magnetic resonance spectroscopy.[7]

A different study by Saven and Lehmann applied a program called SCADS (Statistical Computationally Assisted Design Strategy) to develop a four-helix bundle

protein, PA_{tet}, that selectively binds to a nonbiological cofactor. Similar to Mayo and Dahiyat's work, this program took into account the target structure, energy functions quantifying sequence-structure compatibility, and energetic restraints. SCADS was able to provide the lowest energy conformation based on the backbone, and this method successfully produced a stable, fully asymmetric protein that demonstrated binding specificity.[8] SCADS was later used by DeGrado et al. to successfully design another single-chain four-helix bundle binding a di-iron cofactor, starting from the crystal structure of DF1.[9]

Experimentally, the inverse folding problem can be probed by methods of site-directed mutagenesis. By employing principles of rational design, combinatorial libraries can be synthesized to make large numbers of variants, providing a large database of sequences for further analysis. Through the use of DNA oligonucleotide synthesis and different PCR-based technologies, genes with random mutations can be mass-produced at essentially any position.[3] Previous research has proven that *de novo* libraries can be created with large numbers of variants exhibiting native-like folds and stabilities.[10]

The simplest way to carry out this large-scale analysis is to relate genotype to phenotype through functional selection or an activity screen. The behavior of well-folded proteins typically differs from that of those that are unstructured, and selecting for active variants usually yields sequences adopting native-like folds.[3] Phage display, ribosome display, RNA display and plasmid display libraries are all valuable methods of large-scale selection that facilitate the interpretation of thousands of protein sequences.[11] Screening methods such as these enable the identification of sequences

that form stable folds as well as allow for a better understanding of the sequence-stability relationship that has historically been such a mystery.

1.3 Rop model studies

Repressor of primer, or Rop, is a 63-residue anti-parallel homodimer that binds RNA and regulates ColE1 plasmid replication in *E. coli.*[12] It is an exemplary candidate for site-directed mutagenesis due to its small size and simplicity, and it has been well characterized.[13] It has been crystallized under various conditions, and the high-resolution structure was solved and published by Banner et al. in 1987.[14,15] This alpha-helical protein has been the basis of many model studies in protein engineering, and some of the earliest mutagenesis experiments on Rop were conducted to determine its RNA binding site.[13] Another experiment produced a Rop variant with similar stability and *in vivo* activity to the wild type by changing the two buried cysteine residues to alanine and valine, respectively. This opened the door to a variety of mutagenesis experiments while alleviating the oxidative problems that Cys-proteins are often prone to.[16]

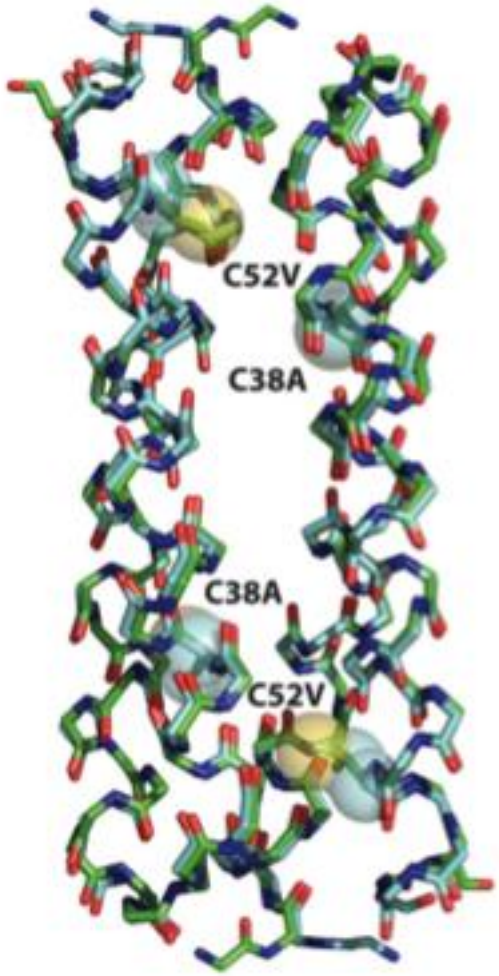


Figure 1.1. Crystal structure of AV Rop, showing the C38A/C52V mutations[16]

Additionally, a cell-based screen has been developed for Rop's function, using Green Fluorescent Protein as a reporter of activity. By cloning GFP into a plasmid with a ColE1 origin, the expression levels of GFP are indicative of Rop function. Presumably, a dim colony would represent an active variant based on the minimal GFP expression, but there are a number of problems associated with having a negative screen for positive hits. The authors discovered that by expressing Rop from

an arabinose promoter and supplementing the LB agar with 0.0005% arabinose, the phenotype reverses to a positive screen, causing fluorescence under conditions of high activity.[17] Regardless of the mechanism, this screen allows for the selection of active variants from hundreds of mutants and has proven to be a formidable asset in large-scale protein sequence analysis.

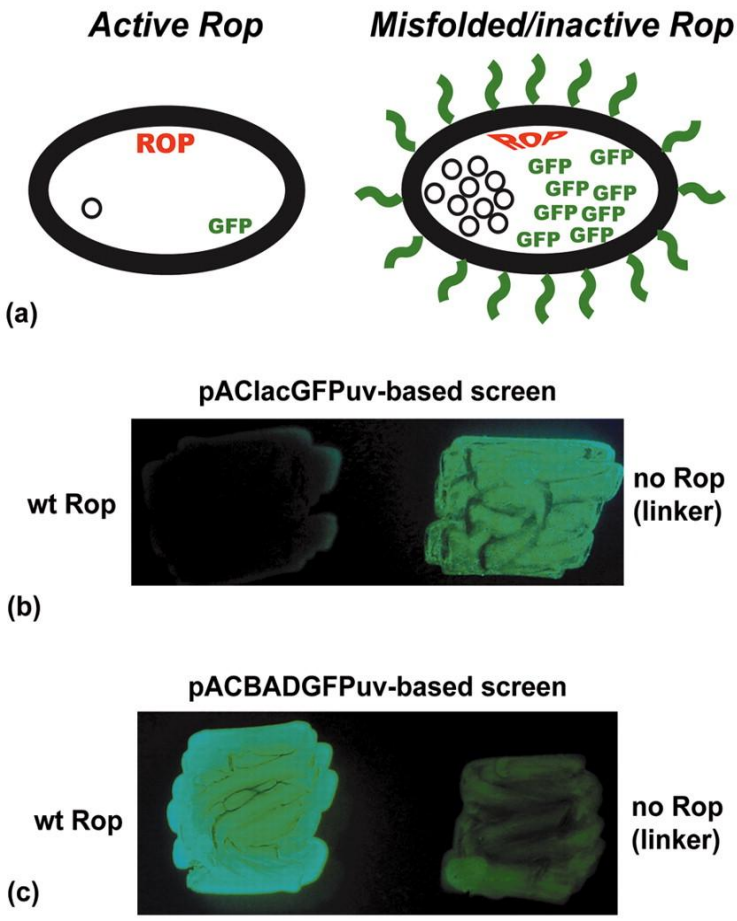


Figure 1.2. a) A pictorial representation of Rop controlling GFP expression levels in the *in vivo* screen. b) The phenotypes of wild type Rop and linker on the negative screen. c) The phenotypes of wild type Rop and linker on the positive screen[17]

1.4 Single-chain Rop

The conversion of Rop to a single-chain monomer is an idea that has been a subject of interest for many researchers in the field. The potential to make asymmetric mutations on a stable model protein is undoubtedly attractive, and mutagenesis studies could be greatly simplified by using a monomer as the model instead of a dimer. Other researchers have been successful in converting other alpha-helical proteins to monomers, and this feat is widely represented in the literature. Lim et al. were able to engineer a monomeric streptavidin (originally a tetramer) through rational design that maintained its binding activity, and they later used this stable monomer to construct dimeric versions of the same protein. Their efforts verify the ability to rationally design oligomeric mutants that are functionally active and stable.[18] Mehl et al. investigated the stability of the dimer interface of an anti-parallel homodimer, GrpE, by making internal deletion mutations to the core. At 12 internal deletion mutations, the dimeric form no longer prevailed, and single chain variants were formed that maintained alpha-helical in structure. Further structural and stability studies have yet to be completed, but it is useful methodology to consider when conducting oligomeric conversions.[19]

A number of single-chain Rop variants have been engineered in the past by experimenting with changes in helix connectivity. The first Rop monomer was created by the Sander lab in 1994, and the success of the design was confirmed by NMR.[20] Regan and Predki continued the production of monomeric Rop with their loop studies in 1995. By redesigning the topology and adding polyglycine linkers, a handful of Rop monomers were designed that were sufficiently similar in structure to

the wild type. The connectivity of these variants is depicted in Figure 3. The initial construct was named Rop_{S33}, with the "S" representing single-chain, the first 3 representing the 3 glycines in loop 1 (connecting 1-1') and the second 3 indicating that there were 3 glycines in loop 2 (connecting 2-2'). Other variants created included Rop_{S34}, Rop_{S43}, Rop_{S44}, Rop_{S45}, Rop_{S54}, and Rop_{S55}. Only Rop_{S44}, Rop_{S45}, Rop_{S54}, Rop_{S55} existed as true monomers (determined by gel filtration).[21]

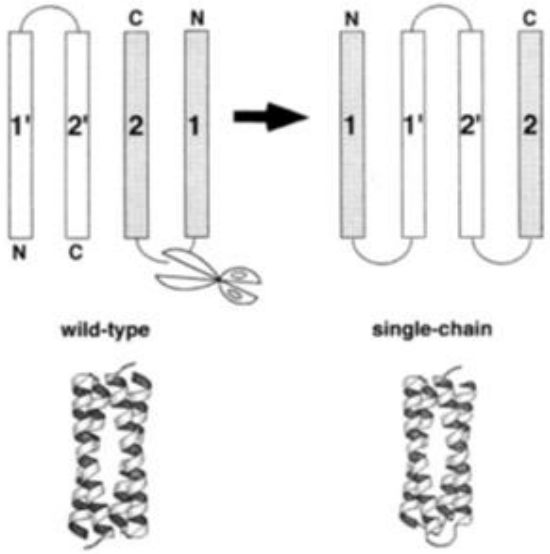


Figure 1.3. A picture showing the redesigned topology of single-chain Rop variants[21]

Kresse et al. were also curious about the design of monomeric versions of Rop. They recognized that there were principally four ways to engineer single-chain variants assuming that the three connective loops are adjacent (Figure 1.4), and they constructed seven different monomers with connective variance. Of the seven, four

were left-handed, and these were utilized to introduce biologically interesting loops with differing amino acid sequences. When expressed in *E. coli*, five out of seven were deemed biologically stable, demonstrating that actual amino acids used in the loops are important in addition to length optimization.[22]

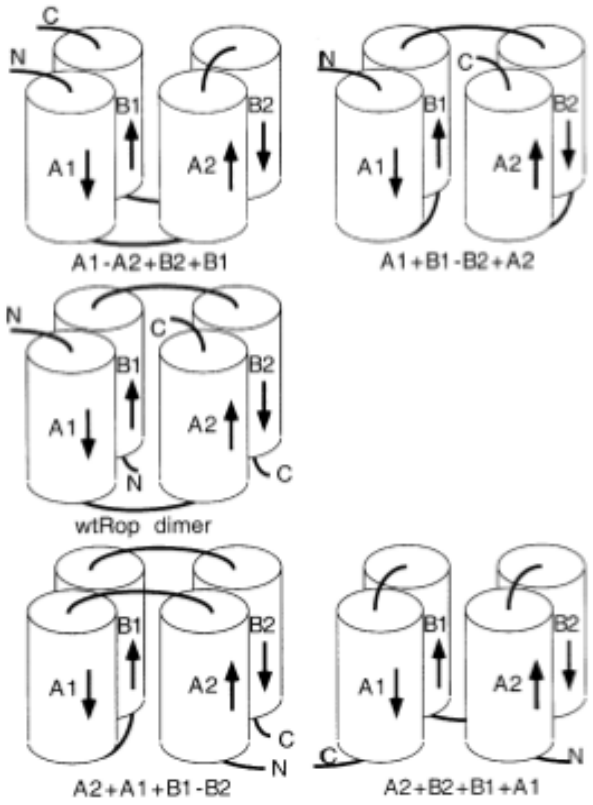


Figure 1.4. Diagrams showing the possible connectivities of single-chain Rop variants[22]

Another version of single chain Rop was devised in 2008 containing a unique buried tryptophan for the purpose of conducting redox studies. This required substantial remodeling of the hydrophobic core in order to shift from parallel to anti-

parallel topology, in addition to an artificial loop connecting the two subunits (Figure 1.5). Termed α_4W to reflect its helical nature and tryptophan mutation, this protein exhibited structural and stability characteristics comparable to the wild type.[23]

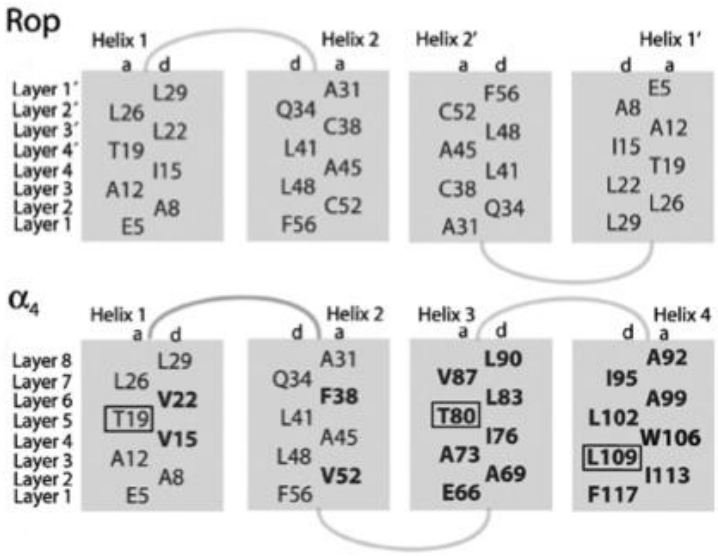


Figure 1.5. Images of the differences in connectivity between wild type Rop and engineered single-chain α_4 variant[23]

If a stable, well-folded Rop monomer could be used as a library scaffold, the collection of data regarding the effects of asymmetric point mutations could be expedited by a combinatorial approach. Other single-chain, four-helix bundle libraries have been reported in the past, and the Hecht lab at Princeton University has worked extensively with variants such as these. Through the binary patterning of hydrophobic and hydrophilic residues, his lab sought to create a multitude of folded yet unselected sequences. "Unselected" in this case means proteins that are *de novo* and have not

evolved through biological or environmental factors. The structure of the most stable protein from their first library, S-824, was resolved by nuclear magnetic resonance and was in accord with their target design.[24] A second generation library was cloned as a follow-up experiment, which yielded another slew of stable, four-helix bundles. The most stable variant from this library was determined by NMR and also folds into the expected four-helix bundle. These studies show that highly stable and well-ordered structures can also occur in unselected libraries, and these monomeric bundle proteins may possess certain qualities analogous to ancestral proteins.[25]

1.5 Loop studies

It is widely believed that the major determinant of protein stability is the packing of hydrophobic regions, whereas peripheral areas can tolerate broader ranges of mutations with minimal consequences. However, the importance of turn and loop regions is often debated in the field of protein folding, and their role in specifying final folded structure is not completely known. Are they simply connectors? If not, how much of an effect do they actually have on overall protein stability?

Jung and Pluckthun were able to improve *in vivo* folding and stability of a single-chain Fv antibody fragment by methods of loop grafting. They discovered that complementary determining regions (CDRs) of antibodies could be grafted onto superior frameworks, resulting in variants that are less prone to aggregation and exhibit much higher thermodynamic stability. Libraries of grafted loop antibodies were screened for stability and binding affinity by phage display, and the results suggested that antibody framework was a major determinant in the physicochemical

properties of the resulting scFv fragment.[26] Helms and Wetzel also found evidence supporting the significance of loop optimization, through very opposite results. Replacement of CDR loops of an immunoglobulin V_L domain proved to be destabilizing to the native structure of the domain, and similar trends were observed in two separate CDRs. The weakening stability of the domain was also seen across a broad scope of mutations, ranging from single point mutations to large insertions.[27] Loops have also been found to have important roles in high affinity binding. Functional analysis of a streptavidin variant showed a mutation to the periphery, not in the active site, was responsible for about a seven-fold improvement in the off rate of the protein.[28]

Multiple studies have also been conducted exploring the effects of loop mutagenesis in Rop. In 1996, Predki, Agrawal, Brunger, and Regan mutated Asp 30, the first residue of the turn region in Rop, to each of the 19 other amino acids. They found that each of the position 30 variants had binding affinities similar to the wild type, and that based on activity alone, this residue could tolerate the substitution of any amino acid. However, through T_m determination by thermal denaturation, it was observed that only 12 of the mutants had improved thermostability when compared to the wild type. These results suggest that there is a distinct preference for certain amino acids at this position on the loop.[29]

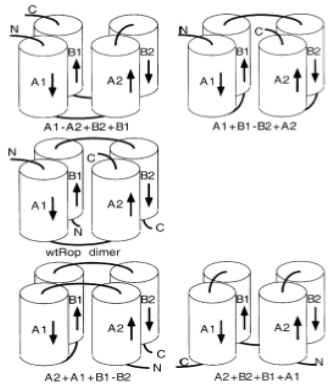


Fig. 1. Schematic representation of the four possible topologies for monomeric Rop variants. Wild-type Rop (wtRop) is shown in the middle for reference. The top two drawings are schematic representations of LM-Rop and RM-Rop. The two left-handed monomers are shown at the left and the right-handed monomers at the right. The helices are shown as cylinders and loop regions as solid lines. The termini are indicated by the letters N and C, respectively. The resulting order of the helices is indicated under the drawings; + signs indicate newly constructed loops and – signs indicate loops that are the same as in wtRop.

Figure 1.6. This shows a closeup of the turn region in Rop. The Asp residue at position 30 is shown in red. The neighboring residues L29 and A31 are shown in yellow.[29]

Regan revisited the implementation of polyglycine linkers with Paul Predki in 1997 when they created Rop variants differing in loop length and tested the resulting stability and rates of protein folding. Native loops were replaced with polyglycine insertions (up to ten residues), and it was determined that the substitution of Gly-Gly led to an increase in flexibility and folding rate. However, with each glycine addition thereafter, the folding rates decreased, indicating a correlation between longer loops and destabilization. They were able to conclude that larger loops in Rop are associated with a greater entropic cost, and that the flexibility of Gly-Gly most likely lowered the energy barrier of the transition state for folding and unfolding.[30]

Thermodynamic parameters for the Gly loop length mutants.

Protein	MRE*	T_m† (°C)	$\Delta G_{56.4}$‡ (kcal mol^{-1})	C_m§ (M)	m_G# (kcal mol^{-1} K^{-1})	ΔG^{0**} (kcal mol^{-1})	$\Delta\Delta G$†† (kcal mol^{-1})
Gly_1	30000	72.7	4.1	2.7	2.0	5.0	1.1
Gly_2	29500	69.0	3.3	2.4	2.7	6.4	2.5
Gly_3	22000	61.2	1.6	2.1	2.8	5.8	1.9
Gly_4	28600	60.3	1.5	2.0	2.5	5.2	1.3
Gly_5	27500	58.4	1.1	1.9	2.6	4.9	1.0
Gly_6	25500	56.7	0.9	1.6	2.6	4.6	0.7
Gly_7	24600	56.3	0.7	1.5	2.4	4.5	0.6
Gly_8	25000	53.8	0.2	1.5	3.1	4.5	0.6
Gly_9	24500	52.9	−0.2	1.4	3.0	4.4	0.5
Gly_{10}	24100	49.8	−0.9	1.3	3.0	3.9	0.0

Figure 1.7. A table of thermodynamic parameters for the Gly loop length mutants[30]

This same family of polyglycine mutants was further studied by Nagi, Anderson, and Regan in 1999. They observed the folding kinetics of these Rop variants using stopped-flow fluorescence, and the signal was observed from the Tyr 49 residue in each monomer. The authors discovered that each of the loop-length mutants exhibited kinetic behavior that was similar to the wild type, showing biphasic refolding kinetics at low concentrations of GuHCl and monophasic as the concentration of the denaturant increased (Figure 1.8). Figure 1.9 shows a graphic comparison of three of the loop length variants as the concentration of GuHCl was increased. The results indicate that the stability of the intermediate is highly dependent upon the loop length, and that as loop length increased, stability of the intermediate decreased.[31]

Protein	k_f (fast phase; s^{-1})	k_f (slow phase; s^{-1})
Gly_2	580	31
Gly_3	190	14
Gly_4	96	11
Gly_5	44	3
Gly_6	45	2
Gly_7	35	2
Gly_8	29	2
Gly_9	17	1.7
Wild-type	29	0.5

Figure 1.8. Refolding rates of the loop-length variants[31]

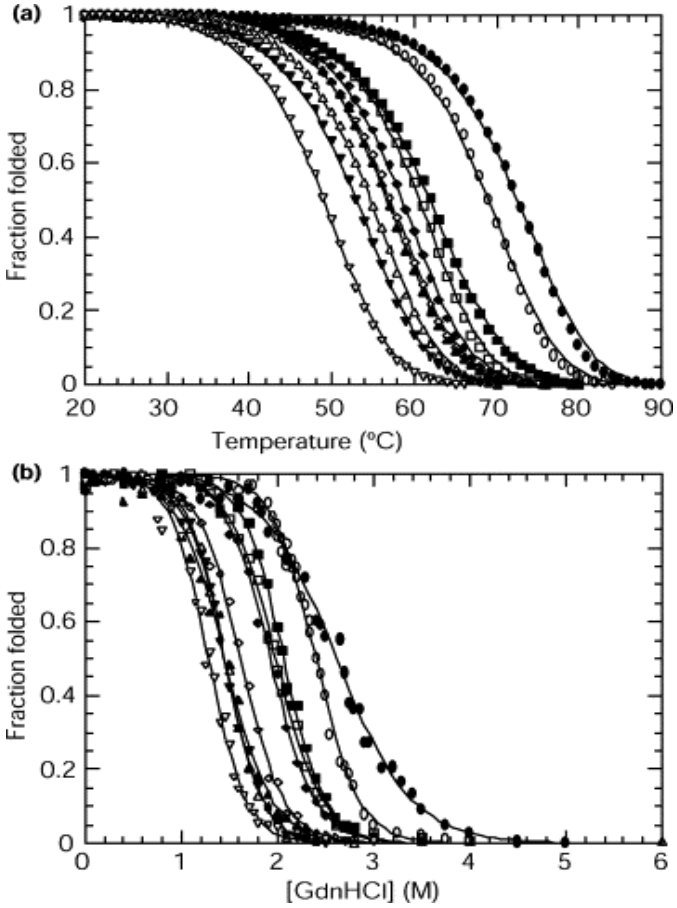


Figure 1.9. (a) Thermal denaturation curves and (b) GdnHCl denaturation curves of the 10 Gly loop mutants, as monitored by CD. Lines in (a) are smoothing fits to the data, while those in (b) are nonlinear least-squares fits to equation 1. Closed circles represent Gly_1; open circles, Gly_2; closed squares, Gly_3; open squares, Gly_4; closed diamonds, Gly_5; open diamonds, Gly_6; closed triangles, Gly_7; open triangles, Gly_8; closed inverted triangles, Gly_9; open inverted triangles, Gly_{10}[31]

All of these studies provide valid evidence that loop length and amino acid identities play active roles in governing protein stability and activity, and that protein sequence in loop regions is not necessarily of secondary importance.

1.6 Surface Mutations

In 1959, Kauzmann suggested the importance of hydrophobic interactions in protein folding. Without the luxury of modern mutational methods or a database of known protein structures, he was a pioneer in exploring the significance of the hydrophobic effect in determining protein structure.[32] Over thirty years later, Ken Dill re-examined this hypothesis and proved that hydrophobicity is in fact the dominant force that drives protein folding.[33] There are only a minimal number of ways a protein can fold to maximize nonpolar interactions in the core, which is potentially why primary sequence only encodes one native structure.

Due to the abundance of evidence indicating the imperative effect of hydrophobicity on protein structure, the significance of other intermolecular forces is questionable. It is widely believed that most electrostatic interactions with the solvent should essentially be indistinguishable in the native and denatured state. In addition, charge-charge interactions on the surface are arguably weakened by high dielectric of solvent, potentially reducing the effects of solvent-exposed residues.[34] Studies have shown that introducing highly charged residues to the periphery can be complicated, and adds little in terms of protein stability. Loss of the conformational entropy of side chains and great desolvation penalties have also been reported.[35] Further insight to this question has surfaced in recent studies, however, and it is

becoming more accepted that solvent-exposed residues can very much affect the outcome of protein stability.

Past authors have studied the contributions of salt bridges and their ability to stabilize proteins. An early study on T4 lysozyme, proved that a partially solvent-exposed salt bridge confers 3-5 kcal/mol to its stability.[36] A different experiment on barstar, the inhibitor of ribonuclease barnase, showed that the replacement of any of its four acidic residues with alanine dramatically improved the stability.[37] This improved stability suggests that the removal of unfavorable electrostatic interactions should be considered in protein engineering. This trend was also observed by Koide's lab when analyzing a type III domain of human fibronectin. Resolving a repulsive electrostatic interaction from a surface Asp 7 located between Asp 23 and Glu 9 through mutagenesis to Asn and Lys, respectively, greatly improved the stability of the protein.[38] Further support of this phenomenon was reported by Spector et al., where the authors used continuum methods to calculate electrostatic contributions of charged and polar side chains to the stability of a small helical protein.[39]

Another approach involves calculating the energy of charge-charge interactions through surface accessibility-corrected Tanford-Kirkwood formalism (TK-SA), as previously conducted by Ibarra-Molero et al.[40] This method can provide reasonable approximations of interactions between amino acids in solvent-exposed regions. Makhatadze et al. utilized this technique to study two oppositely charged residues on the surface of ubiquitin, K11 and E34. These amino acids are spatially located in a way that is characteristic of salt bridges, and a "double mutant" analysis was performed to test the strength of the interactions. Results showed that the salt

bridge of reverse orientation (E11/K34) was similar in strength to the original pair, but that the global stability of K34/E11 was higher by 2.2 kJ/mol. The difference in overall stability of ubiquitin between these two variants indicates that the charge-charge interactions between the residues forming the salt bridge are a major contributor. Thus, while salt bridges have been shown to be stabilizing, the positive effects are largely dependent on the electrostatic interactions of surface residues. Analysis of the energetics of the possible 16 salt bridges from six unique proteins further confirmed this conclusion.[41]

Comparison of stability differences between thermophiles and mesophiles is another interesting way to study the effects of surface mutations on thermostability. Two cold shock proteins differing at 12 positions, one of mesophilic and one of thermophilic nature, were the subject of an analysis conducted by Perl et al. The authors reported that the improvement of stability by 15.8 kJ/mol in the thermophilic protein was attributed to electrostatic contributions of two surface residues. Additionally, high thermostability could be achieved by the mesophilic variant after the subsitution of the Glu residues at positions 3 and 66 to Arg and Leu, respectively.[42] Clearly, protein stability can be greatly enhanced or decreased by site-directed mutagenesis of solvent-exposed residues, and should be considered in applications of protein engineering.

Chapter 2

Objectives

2.1 Synthesis and characterization of single-chain Rop variants

As previously discussed, wild type Rop exists as a 63-residue antiparallel homodimer. Through the redesign of loop connectivity, dimeric versions of Rop can be converted to monomers. While previous researchers have engineered single-chain Rop variants, none of them were specifically designed to mimic the stability and activity of the wild type. The goals for the target variants to be created include being functionally active and cysteine-free, and having similar stability to wild-type Rop.

2.2 Biophysical characterization of Rop surface variants

Due to the controversy surrounding the importance of solvent-exposed residues in protein stability, a NNK_5 library mutating 5 of the most solvent-exposed positions of Rop was created. This library was constructed and screened by previous masters student Chau Nguyen and current graduate student Shiladitya Sen. They also screened the library using the activity screen described above and sequenced the positive hits. The goal of this project is to biophysically characterize these active surface variants and to draw conclusions about the effects of surface mutations on overall protein stability.

Chapter 3

Materials and Methods

3.1 Synthesis of single-chain Rop variants

Gene construction

Two sets of oligonucleotides were purchased from Sigma-Genosys to synthesize genes with varying loop mutations using PDB ID: 1yo7 as the scaffold. The first variant, with isoleucine, was designed to have a point mutation at position 57 (Lys/Gly) and an insertion (Ser) in between the glycines at positions 88 and 89. The second variant, without isoleucine, contained the same point mutation and insertion in addition to an Ile deletion at position 62. The primers used to generate these mutants are listed below in Table 3.1. Methods of overlap PCR were used for the reassembly, and the reactions were performed on a Flexigene thermal cycler (Techne Inc). The genes were assembled from 5 different primers, amplified in two separate reactions: one with the first two oligonucleotides, and the other with the remaining three. Amplification reactions were typically carried out at 50 μL volumes at 57-60 °C for 25 cycles. Pfu polymerase used was purified in-house. Each of these contained a BsaI site (primers shown in Table 1), and they were digested with BsaI (NEB) at 37 °C for 8 hours. These two reactions were then ligated overnight at 16 °C. All ligations were cleaned up using a Qiagen PCR purification kit and protocol.

1yo7-1	ACTAAGCAAGAGAAGACAGCACTTAATATGGCTCGTTTTATTCGTTCTCAAACTCTTACTCTTCTTGAAAAACTTAATGAACTTGATGCTGACG
1yo7-2	gAAAGAAGCAAGAACAGAACGATAAAGTTCATCAGCATGATCATGtaaAGATTCAGCAATATCAGCTTGTTCGTCAGCATCAAGTTCATTAAG
1yo7-2nd-a	GATGAaCAgGCtGACATcGCAGAAAGCttaCATGACCATGCAGAtGAACTGTATCGTAGTGTGCTGGC
1yo7-2nd-b	GCCATGTTCAGTGCTGTTTTTTCTTGTTTAGAACCGCTACCAAAaCGAGCCAGCACACTACGATACAG
1yo7-2nd-c	AAAAACAGCACTGAACATGGCACGTTTtATTCGCAGTCAAACCCTGACCcttCTGGAAAAATTAAATGAGCTT
bsa1-2prime	ATTATTGGTCTCAAcGGTCAAGATGAaCAgGCtGACATc
bsa1-2	AATTATGGTCTCACCGTTTTTACCGAAAGAAGCAAGAACAGAACG

Table 3.1. Primers used to construct single-chain variants

Cloning

Each variant was amplified to form genes used for screening in the *in vivo* functional assay and expression genes containing N-terminal TEV and His_6 sites. The screening genes and pACT7lac vector (constructed by Magliery et al) were digested with AflIII and BamHI (NEB) for 2 hours at 37 °C, and the expression genes and pMRH6 vector (constructed by Magliery et al) were digested with AflII and BanI (NEB), again for 2 hours at 37 °C. The genes were ligated into the vectors overnight at 16 °C and transformed into electrocompetent DH10B. The recovered cells were plated on semisolid LB agar with 30 μg/mL kanamycin and incubated overnight at 37 °C. Colonies were picked and DNA was extracted (Qiagen Miniprep) and sequenced (Genewiz).

Screening

Single-chain variants in pAC vectors were transformed into electrocompetent DH10B (Stratagene), which already contained the screening plasmid pUCBAD-GFPUV (constructed by Magliery et al.) Cultures were plated onto a semisolid agar plate containing 30 μg/mL kanamycin, 100 μg/mL ampicillin, and 0.0005% (w/v) arabinose and incubated overnight at 42 °C. Plates were analyzed under UV lights, and different levels of fluorescence were observed.

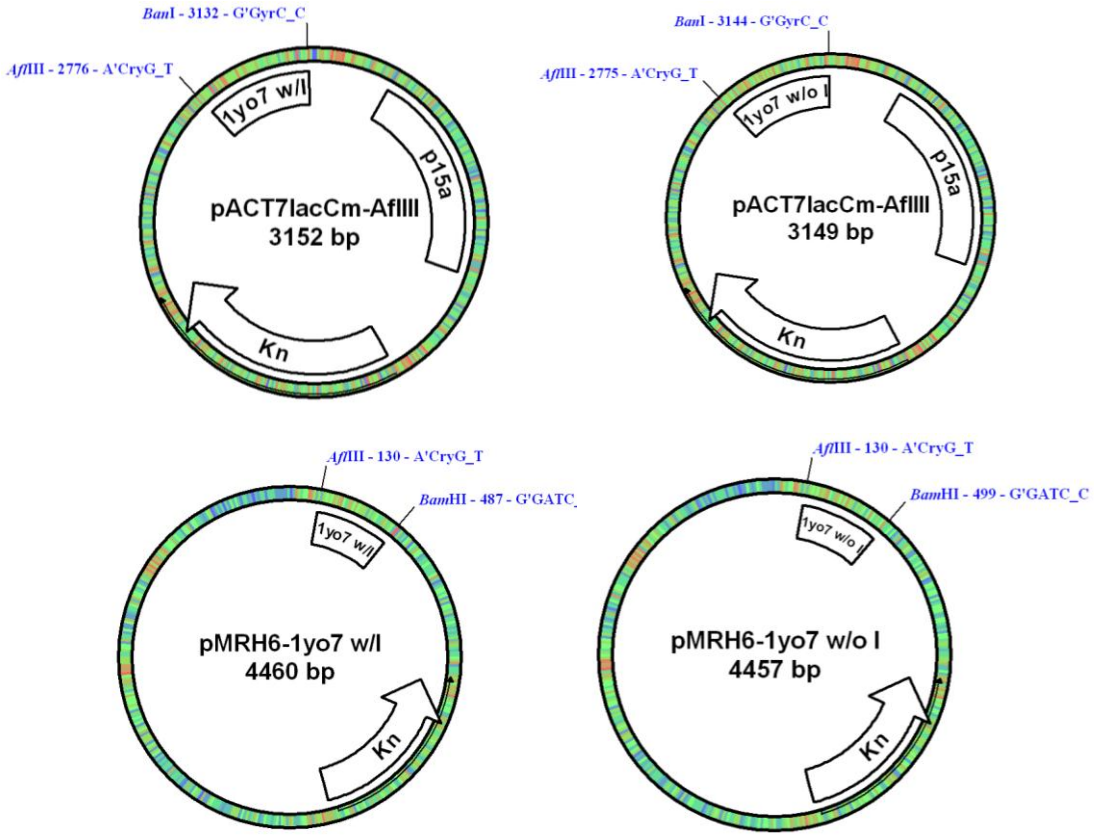


Figure 3.1. Plasmid maps of single-chain variants in screening and expression vectors

Protein purification

Single-chain variants in pMR vectors were transformed into electrocompetent C41(DE3) cells (Stratagene), and 1 L cultures were grown in 2YT media. The cultures were induced with 0.1 mM IPTG during the log growth phase, and expression occurred overnight at 30 °C or at 37 °C for 3-4 hours. Harvested cells were frozen at -80 °C, and pellets were resuspended in 25 mL lysis buffer (50 mM Tris-HCl, 300 mM NaCl, 10 mM imidazole, pH 7.4) mixed with 5 mM $MgCl_2$, 0.5 mM $CaCl_2$, 5 μL DNaseI, 10 μL RNase (10 mg/mL stock), 0.1% Triton X-100, and 0.1 mM PMSF protease inhibitor. Cell lysis was achieved by sonication (Misonix Ultrasonic Liquid Processors, Model Q125) at half power for 6 x 30 seconds, with incubation on ice for 2 minutes in between pulses. Clear cell lysate was then separated from the pellet through centrifugation and mixed with 500 μL of Ni-NTA slurry (Qiagen) on ice for at least one hour. The mixture was poured through a pre-fritted column, and the flowthrough was allowed to go to waste. The protein was then washed with 18 mL of wash buffer (50 mM Tris-HCl, 300 mM NaCl, 20 mM imidazole, pH 7.4) and eluted with 3 mL of elution buffer (50 mM Tris-HCl, 300 mM NaCl, 250 mM imidazole, pH 7.4). TEV protease (0.5 mg) was added twice and allowed to incubate at room temperature overnight or 30 °C for 3 hours. The solution was desalted with PD-10 columns (GE Amersham) into lysis buffer and mixed with 500 μL Ni-NTA slurry (Qiagen) on ice for a minimum of one hour. The solution was added to a pre-fritted column again and cleaved protein was collected. Protein was concentrated through a YM-3 filter (Millipore) and dialyzed into appropriate buffers for experimental analysis.

CD spectroscopy

Both variants were analyzed on a Jasco J-185 CD Spectrometer at 25 μM concentrations, determined by UV absorption at 280 nm. The variants were exchanged into CD buffer (50 mM sodium phosphate, 300 mM NaCl, pH 6.3) and wavelength scans were taken from 190-260 nm. Thermal denaturations were taken at 1 °C/min from 25-90 °C at 222 nm. Samples at the same concentration and buffer conditions were denatured by varying concentrations of urea up to 8 M and allowed to equilibrate for at least 24 hours at room temperature. Scans were also taken at 222 nm.

HSQC-NMR

Each variant was purified using the aforementioned protocol, but in minimal media instead of rich 2YT. ^{15}N-labeled NH_4Cl was used as the nitrogen source, and variants were exchanged into CD buffer by dialysis. Samples were concentrated to 1 mg/mL and mixed with 10% D_2O directly before experiments were conducted. ^{1}H- and ^{15}N-HSQC spectra were obtained from a 600 MHz Bruker NMR with inverse probe using standard Lewis E. Kay pulse sequences at 30°C. The actual scans were performed by graduate student Shiladitya Sen and Chunhua Yuan.

Gel filtration

Gel filtration experiments were conducted on the variants to determine their oligomeric state in solution. Using a Superdex 75 10/300 GL column (GE Tricorn) and phosphate running buffer (CD buffer), each variant was loaded onto the column and the UV absorption of eluted material was recorded. These experiments were conducted using an ÄKTA FPLC at 4 °C at protein concentrations of 0.5 mg/mL,

determined by gel. The variants were exchanged into CD buffer and concentrated prior to loading the column.

Crystallography

Both variants were crystallized using 2.5-15% PEG 6000 precipitant at pH of 5.1-5.4 in 0.1 M acetate with the addition of 20% glycerol and 0.1 M NaCl. Crystals were obtained from sitting drop trays at room temperature with a 2:1 protein to mother liquor ratio. Protein samples were exchanged into CD buffer and concentrated to about 2 mg/mL (determined by gel) prior to setting up trays. Crystals for the variant without isoleucine were sent for diffraction to APS at the Eli Lilly Beamline. A data set was collected and integration was performed with d*TREK. Molecular replacement modeling and building is currently being conducted using the Phenix suite and Phaser.

3.2 Biophysical characterization of Rop surface variants

Library construction

In the NNK_5 surface library, 5 solvent-exposed residues at positions 38, 39, 42, 45, and 45 were randomized to all 20 amino acids. The library was constructed using an engineered Cys-free Rop variant as the scaffold (C38A/C52V) called AV Rop. The library was constructed by reassembly PCR of the two fragments 1yo7-5' and NNK surf-term (Table 3.2) by using Pfu polymerase (produced in house). The reassembly product was PCR amplified using two primers pAC-pro and pAC-term (Table 3.2. This library of inserts was digested with AflIII and BanI and then cloned into the pACT7lacCm-AflIII vector (Figure 2). The background of the original vector

was removed by digestion with EcoRI and NcoI (NEB). This work was done by Master's student Chau Nguyen. Graduate student Shiladitya Sen recloned this library into the pMRH6 vector at the BamHI and AflIII restriction sites (NEB).

1yo7-5'	ACTAAGCAAGAGAAGACAGCACTTAATATGGCTCGTTTTATTCG TTCTCAAACTCTTACTCTTCTTG AAAAACTTAATGAACTTGATGCTGACG
NNKsurf- term	CAAAGCGAGCAAGAACAGAACGATAAAGMNNMNNAGCATGMN NATGTAAMNNMNNAGCAATAT CAGCTTGTTCGTCAGCATCAAGTTCATTAAG
pMR-pro	AATATTACACGTGGTGAGAATCTGTATTTTCAGGGCACTAAGCA AGAGAAGACAGCAC
pMR-term	TAATAAGGATCCTCACAGATTCTCACCATCATCACCAAAGCGAG CAAGAACAGAAC
pAC-pro	AATAATACACGTATGACTAAGCAAGAGAAGACAGCAC
pAC-term	AATAATACACGTATGACTAAGCAAGAGAAGACAGCAC

Table 3.2. NNK_5 surface oligonucleotide sequences

Screening

The NNK_5 surface library was transformed into DH10B containing pUCBAD-GFPUV and grown on a LB plate containing 30 μg/mL kanamycin, 100 μg/mL ampicillin, and 0.0005% (w/v) arabinose overnight at 42 °C in order to screen for *in vivo* activity. This method was employed to distinguish active variants from inactive variants in the surface library. Fluorescent colonies were picked, the DNA was extracted (Qiagen) and the active variants were sequenced (Genewiz). This work was also completed by Chau Nguygen and Shiladitya Sen.

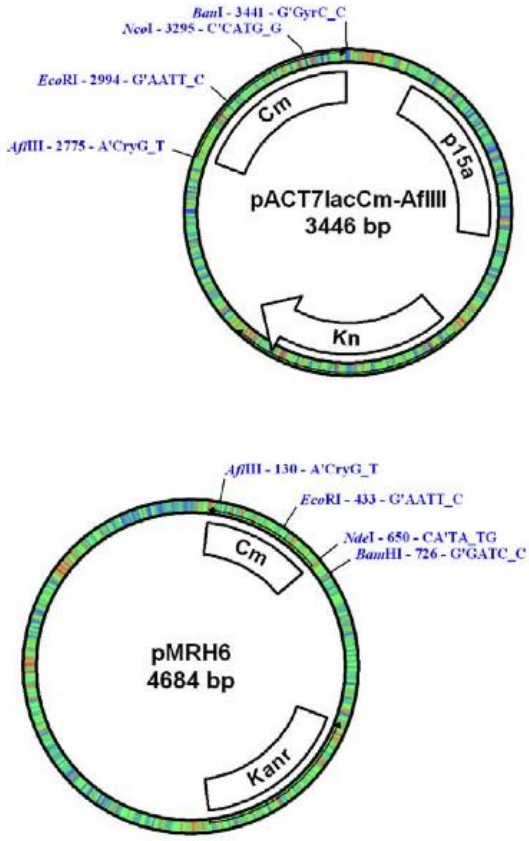


Figure 3.2. Library cloning plasmid maps

Protein purification

Surface variants in pMR vectors were transformed into electrocompetent BL21(DE3) cells (Stratagene), and 1 or 2L cultures were grown in 2YT media. The cultures were induced with 0.1 mM IPTG during the log growth phase, and expression occurred overnight at 30 °C or at 37 °C for 3-4 hours. Harvested cells were frozen at -80 °C, and pellets were resuspended in 25 mL lysis buffer (50 mM

sodium phosphate, 300 mM NaCl, 10 mM imidazole, 5 mM β-mercaptoethanol, pH 7.4) mixed with 5 mM $MgCl_2$, 0.5 mM $CaCl_2$, 5 μL DNaseI, 15 μL RNase (10 mg/mL stock), 0.1% Triton X-100, and 0.1 mM PMSF protease inhibitor. Cell lysis was achieved by sonication (Misonix Ultrasonic Liquid Processors, Model Q125) at half power for 6 x 30 seconds, with incubation on ice for 2 minutes in between pulses. The mixture was allowed to incubate for 30 minutes in a room temperature water bath. Clear cell lysate was then separated from the pellet through centrifugation and mixed with 800 μL of Ni-NTA slurry (Qiagen) on ice for at least one hour. The mixture was poured through a pre-fritted column, and the flowthrough was allowed to go to waste. The protein was then washed with 15 mL of wash buffer (50 mM sodium phosphate, 300 mM NaCl, 20 mM imidazole, 5 mM β-mercaptoethanol, pH 7.4) and eluted with 3 mL of elution buffer (50 mM sodium phosphate, 300 mM NaCl, 250 mM imidazole, 5 mM β-mercaptoethanol, pH 7.4). TEV protease (0.5 mg) was added twice and allowed to incubate at room temperature overnight or 30 °C for 3 hours. Protease inhibitor PMSF was added to 0.02 mM during the second TEV reaction. The solution was desalted with PD-10 columns (GE Amersham) and mixed with 800 μL Ni-NTA slurry (Qiagen) on ice for a minimum of one hour. The solution was added to a pre-fritted column again and cleaved protein was collected. Protein was concentrated through a YM-3 filter (Millipore) and dialyzed into appropriate buffers for experimental analysis.

Due to instability at pH 7.4 and 4 °C, certain variants required alternate methods of purification. Two changes were made for the purification of ARFWE. The entire purification was done at room temperature and eluted protein was dialyzed

into CD buffer (pH 7.4) for immediate biophysical characterization. Only one change was made to the purification of EEDHM, which was purified using buffers containing the addition of 10% glycerol. The CD buffer for characterization did not contain 10% glycerol.

CD spectroscopy

Surface variants were analyzed on a Jasco J-185 CD Spectrometer at 50 μM concentrations, determined by UV absorption at 280 nm. The variants were exchanged into CD buffer (50 mM sodium phosphate, 300 mM NaCl, pH 6.3) and wavelength scans were taken from 190-260 nm. Thermal denaturations were taken at 1 °C/min from 25-90 °C at 222 nm. This work was done in collaboration with Kim Stephany.

HSQC-NMR

Each variant was purified using the aforementioned protocol, but in minimal media instead of rich 2YT. ^{15}N-labeled NH_4Cl was used as the nitrogen source, and variants were exchanged into CD buffer by dialysis. Samples were concentrated to 1 mg/mL and mixed with 10% D_2O directly before experiments were conducted. ^{1}H- and ^{15}N-HSQC spectra were obtained from a 600 MHz Bruker NMR with inverse probe using standard Lewis E. Kay pulse sequences at 30 °C. The actual scans were performed by graduate student Shiladitya Sen and Chunhua Yuan, and purification and preparation of samples was done in collaboration with Kim Stephany.

Gel filtration

Gel filtration experiments were conducted on the variants to determine their oligomeric state in solution. Using a Superdex 75 10/300 GL column (GE Tricorn)

and phosphate running buffer (CD buffer), each variant was loaded onto the column and the UV absorption of eluted material was recorded. These experiments were conducted using an ÄKTA FPLC at 4 °C at protein concentrations of 150 μM determined by UV absorption at 280 nm. This concentration was chosen because it was the minimum required to produce a visible UV signal. The variants were exchanged into CD buffer and concentrated prior to loading the column.

Solubility experiments

Cultures of each variant were inoculated in 5 mL media and grown overnight at 37 °C for a small-scale expression test to check for protein solubility. The cultures were diluted to an optical density of 0.8, grown at 37 °C for 10 minutes and then induced with 0.1 mM IPTG. Cultures were grown at 37 °C for 3 hours, normalized, and collected by centrifugation. The pellets were resuspended in 125 μL of lysis buffer (50 mM sodium phosphate, 300 mM NaCl, 10 mM imidazole, 5mM β-mercaptoethanol, pH 7.4) plus the addition of 5 mM $MgCl_2$, 0.5 mM $CaCl_{2,}$, 2 mg/mL DNase I, and 6 mg/L RNase A. Cells were lysed using glass beads and a vortex. Lysate was decanted and pellets were washed three times with 1 mL of lysis buffer. A volume of 20 μL of SLB loading buffer was added to pellets, which were were heated and vortexed five times. 2 μL of the pellet solution was mixed with another 18 μL of SLB loading buffer, and 10 μL of each were loaded on a 18% SDS-PAGE gel in addition to 10 μL of lysate mixed with 10 μL SLB loading buffer for comparison. This experiment was completed by Kim Stephany.

Activity experiments

Surface variants were transformed into DH10B(DE3) cells already containing the pUCBAD-GFPUV plasmid. An overnight culture of each variant was then grown at 42 °C, cells were normalized and diluted 10^4, and 10 μL of each dilution was plated on a semisolid LB agar plate containing 30 μg/mL kanamycin, 100 μg/mL ampicillin, and 0.0005% (w/v) arabinose. The plate was incubated overnight at 42 °C.

To confirm the fluorescent phenotypes present shown from these results, we compared the copy number of the ColE1 plasmid in AV Rop and an inactive linker to our surface variants by miniprepping (Qiagen) the DH10B(DE3) pUCBAD-GFPuv cells and comparing the plasmid levels. Similar to the first activity screen, overnight cultures of each variant were grown at 42 °C and normalized. Normalized amounts of culture were miniprepped, and an analytical digest at the XbaI and XmaI restriction sites (NEB) was completed. The digests were run on a 1% agarose gel to determine plasmid levels of each variant. This experiment was also carried out by Kim Stephany.

Chapter 4

Discussion and Results

4.1 Synthesis and biophysical characterization of single-chain Rop variants

When choosing the scaffold to use for the construction of new single-chain Rop variants, we decided to pick the best Kresse variant that was crystallized, PDB ID: 1yo7, although there is no paper published with this result. By superimposing 1yo7 with wild type Rop and observing where the proteins differed the most structurally, conclusions on how to rationally design new single-chain variants were made. Two loop sequences were of particular interest: KKNGQI (the artificial linker) and the native loop GGS. Shown below in Figure 4.1 is 1yo7 (in green) superimposed against wild type Rop (in blue) with these loops indicated in red. It can be seen that the KKNGQI loop (left image) seems to form an extra helix-turn, and previous loop studies done on Rop have indicated that a 6-residue linker may not be entropically favorable. Additionally, it appears that the two helices connected by the GGS loop (right image) are being pulled too closely together, and it was thought that perhaps this loop needed an insertion to provide it with more flexibility and relax the apparent tension between the two bundles.

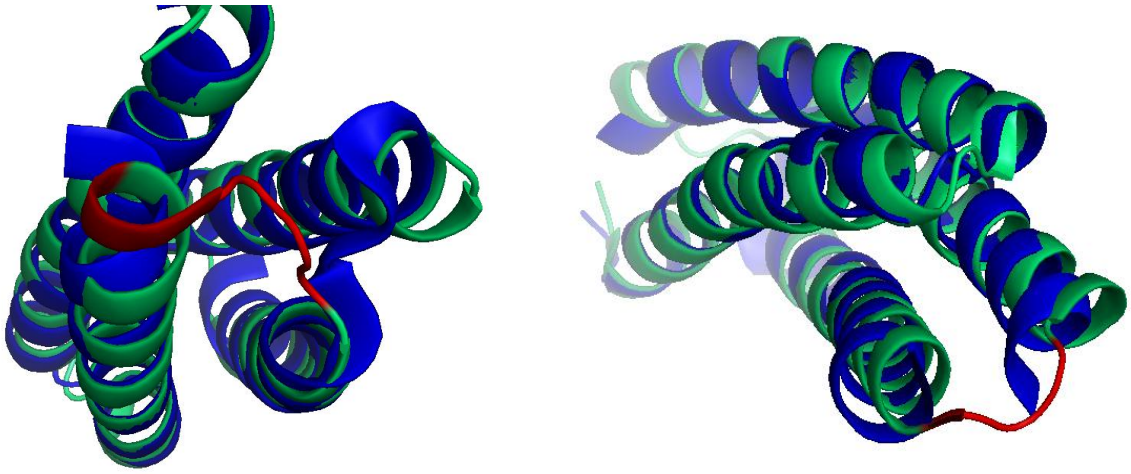


Figure 4.1. Wild type Rop (blue) superimposed against 1yo7 (green) with loops indicated in red

The loop residues comprising the KKNGQI loop are located at positions 57-62 in 1yo7, and the residues that make up the GGS loop are at positions 88-90. A Ramachandran plot of 1yo7, shown in Figure 4.2, indicates that many of these residues fall in places that are unfavorable for alpha-helical proteins. This plot suggests that the loop residues in 1yo7 are not ideal, and that loop modifications are potentially a good place to start when trying to improve the stability and activity.

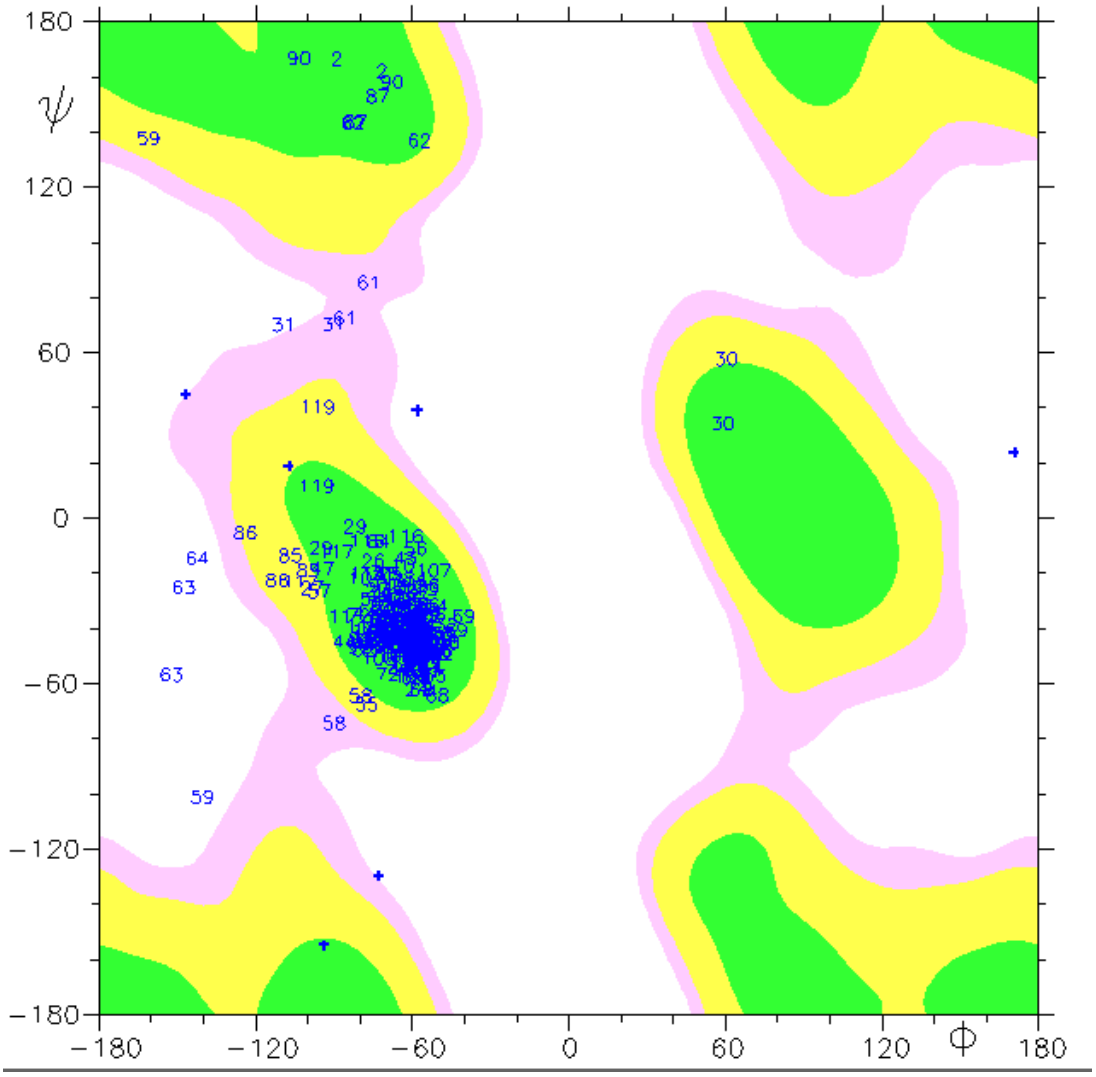


Figure 4.2. A Ramachandran plot of 1yo7

Dr. Jeremy Beck, a former member of the Hadad group from The Ohio State University Department of Chemistry, completed MD simulation studies on single-chain variants of Rop. He constructed the A1B1B2A2 variant of single-chain Rop using the AV Rop double mutant as a base structure. A NSGT linker was added between the A1 and B1 helices utilizing residues 65-69 from PDB ID: 1NZN. A

second FDGD linker was added between the B2 and A2 helices, utilizing residues 99-103 from PDB ID: 1M6E. C-terminal residues 115-126 were omitted from backbone. He found that the only variation between wild type Rop and the A1B1B2A2 single-chain Rop variant exists at the A1B1 loop and B2A2 loops, which are aligned parallel to one another. The structure of A1A2B2B1 and its loop sequences are shown in Figure 4.3.

The 4-helix bundle of A1A2B2B1 is predicted to be more rigid than wild type Rop, and the largest deviation in structure between wild type Rop and A1A2B2B1 occurred at the inserted A2B2 loop and B2B1 loop. In the A2B2 loop, helical structure was lost from D60-A63 due to the constraints of the short loop. During the final 1 ns of simulation time, this region completely lost its alpha-helical character, which corresponds to the increase in backbone RMSD. Similarly, the B2B1 loop lost its alpha-helical character compared to the wild type, with A82 and R83 adopting a random strand backbone orientation. Dr. Beck concluded that this loss of character may have been due to the length of the inserted loop, as the initial searches in Arch-Pred required the addition of the vicinal K88 to the loop searches in order to find a suitable insertion loop. It is also possible that the loop was too short for the B2B1 linker, leading to the observed unwinding of the end of the helical residues.

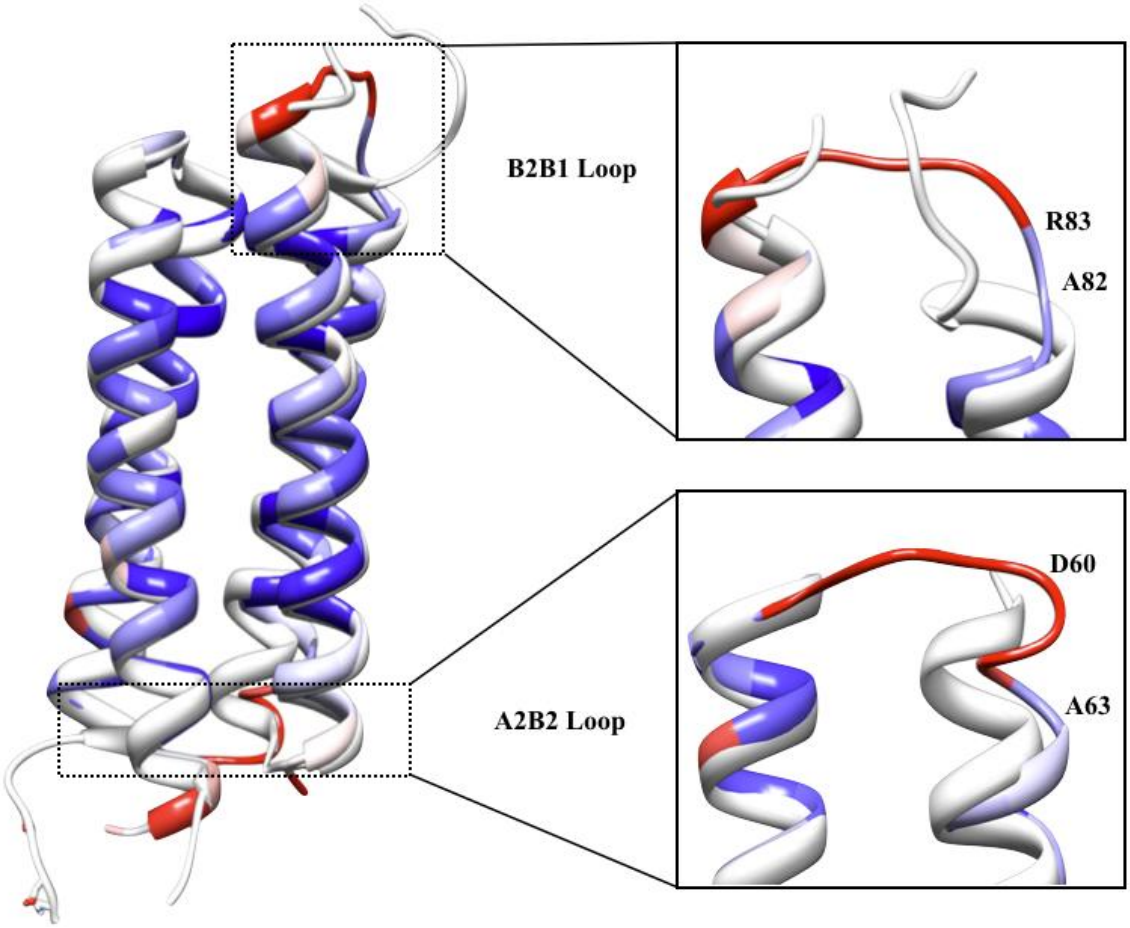


Figure 4.3. MD simulation of the A1A2B2B1 structure. White residues represent a per-reside atomic fluctuation of 1.25 $Å^2$, blue residues are more rigid (dark blue at 0.5 $Å^2$) and red residues are more flexible (dark red at 2.0 $Å^2$). Structures of the inserted A2B2 loop and B2B1 loop are depicted at right, indicating loss of alpha-helical structure from wild type Rop and high flexibility.

All of this data prompted the site-directed mutagenesis of the loop residues of 1yo7 in hopes of engineering new single-chain Rop variants of similar stability and activity to the wild type. Considering the extra turn of the helix given by the KKNGQI loop, the K57 residue was mutated to a Gly, which is a known helix-breaker. Based on previous studies and simulations indicating that loop length was an entropic issue, one variant was constructed with the deletion of I62. It was thought that the removal of this residue would be energetically favorable by shortening the loop, as well as enhance its solubility by the removal of hydrophobic side chains that were previously solvent-exposed. Additionally, both variants were designed with a Ser insertion in between G88 and G89, with hopes of improving flexibility and relieving the tension caused by such a tight turn.

When tested in the *in vivo* activity screen discussed earlier, the two variants showed considerably more fluorescence than 1yo7, indicating improved activity. The results are shown below in Figure 4.4, comparing the variants with and without isoleucine, 1yo7, AV Rop (the positive control that will be used throughout the characterization of these variants), and link, which is a negative control with no Rop gene present. The engineered variants show phenotypes that are significantly more fluorescent than the negative control and 1yo7, and that are arguably as fluorescent as the positive control.

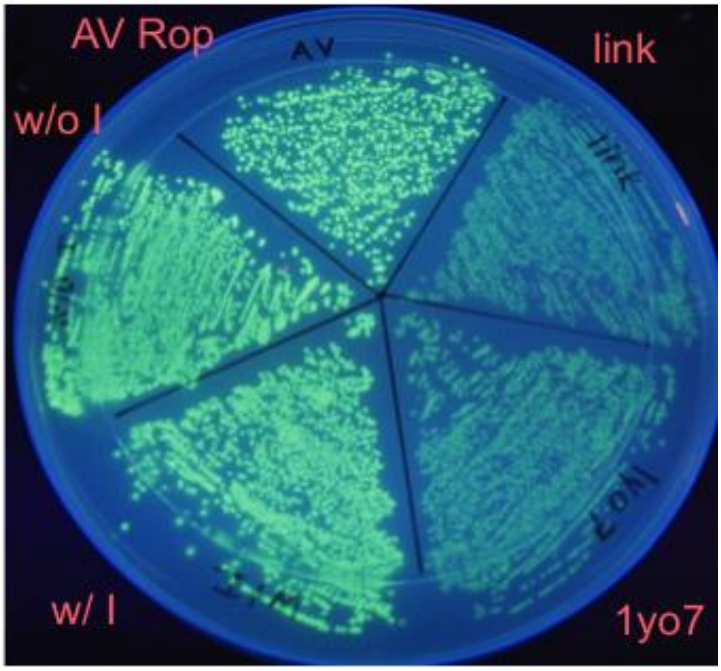


Figure 4.4. Comparison of single-chain variants on GFP activity screen with 1yo7, positive and negative controls

Additionally, these variants were analyzed through circular dichroism. CD scans shown below in Figure 4.5 have peaks at 208 and 222 nm, indicating the alpha-helical nature of the proteins. Thermal denaturation (Figure 4.6) shows cooperative unfolding of both variants, with $T_{1/2}$ values similar to the positive control. It should be noted that neither of the single-chain variants exhibit reversible folding. Urea denaturation produced similar dynamic results, indicating the cooperative unfolding of the variants with and without isoleucine (Figure 4.7). 1yo7 shows limited thermodynamic and dynamic stability when compared to the two engineered variants.

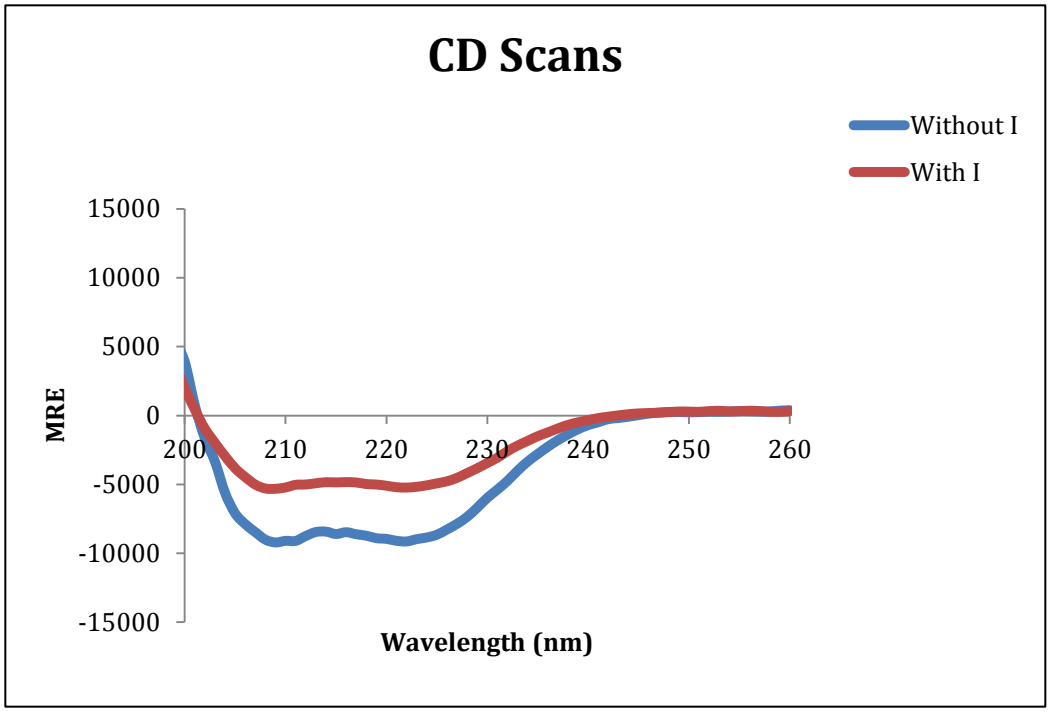


Figure 4.5. CD scans of single-chain variants

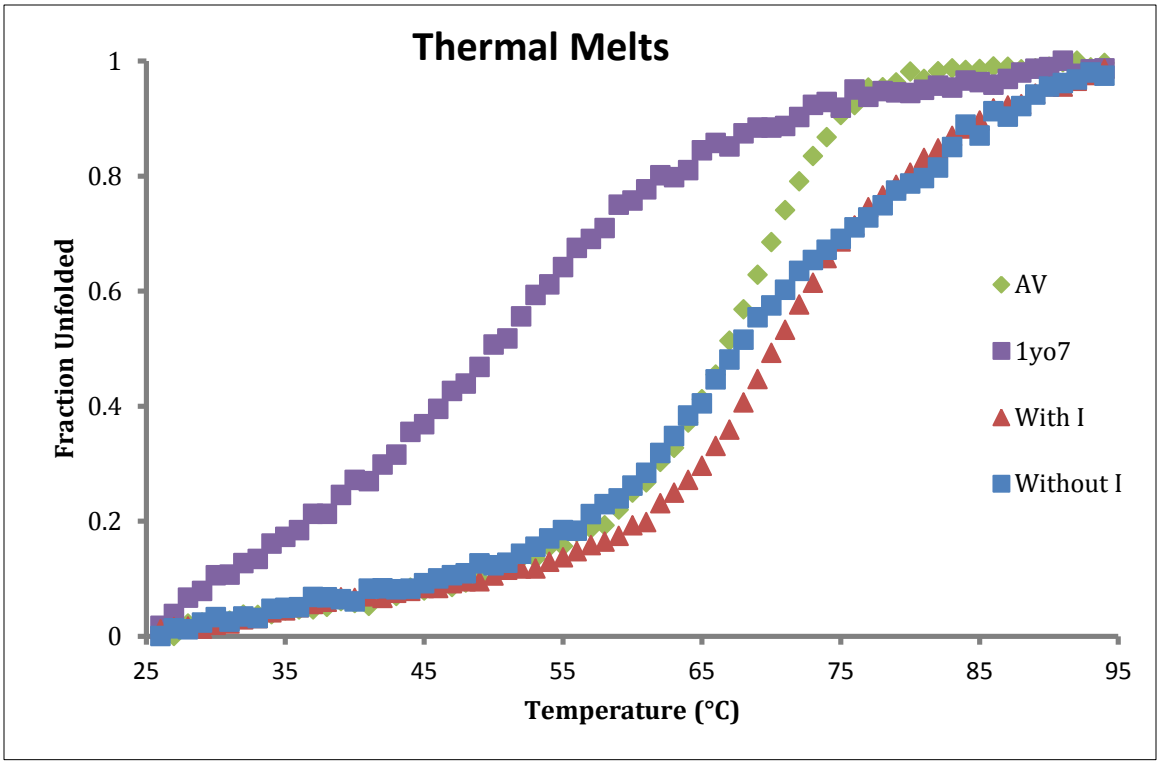


Figure 4.6. Thermal melts of single-chain variants, AV Rop, and 1yo7

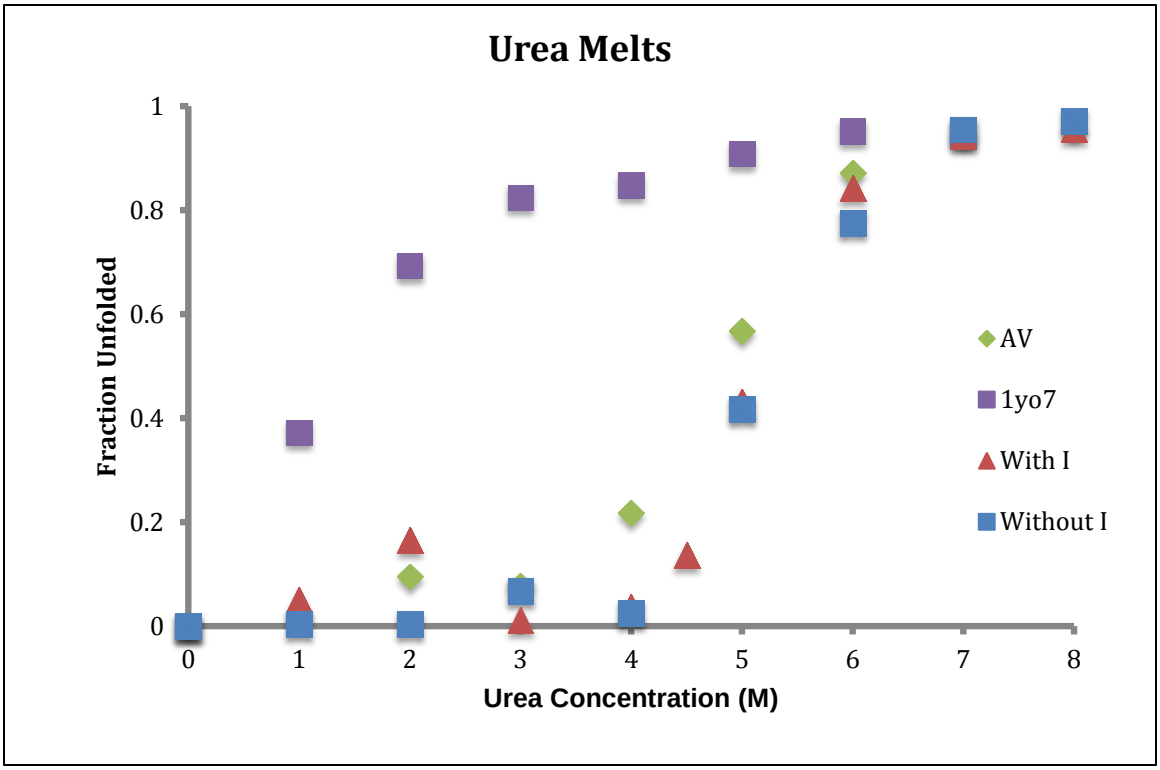


Figure 4.7. Urea melts of single-chain variants, AV Rop, and 1yo7

Additionally, the HSQC spectra for these single-chain variants indicate that the proteins are well-folded. The peaks are widely dispersed, and the fact that there are fewer peaks than amino acid residues may suggest that at least some symmetry was retained (Figure 4.8).

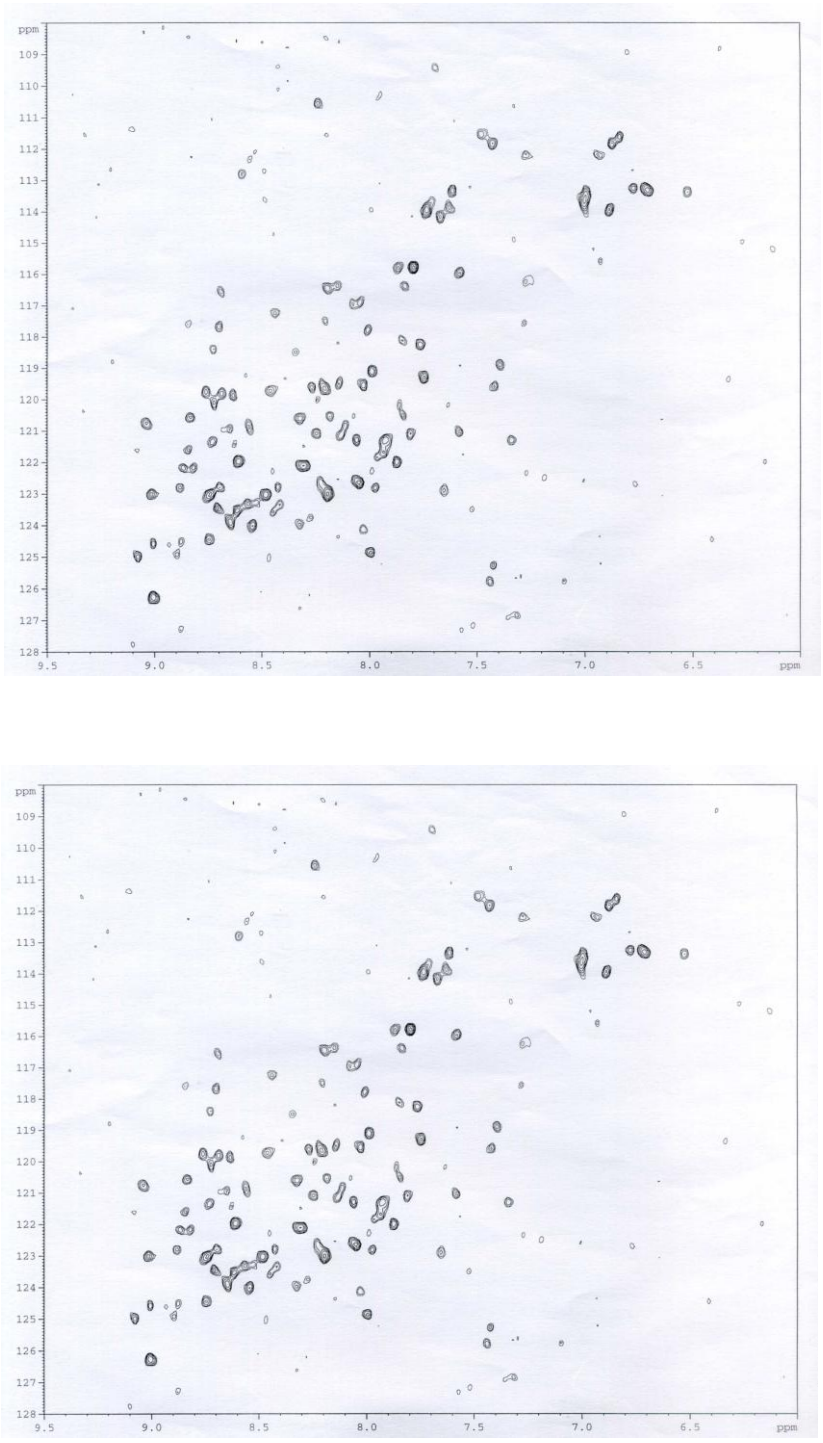


Figure 4.8. HSQC scans of the variant with isolecuine (above) and without isoleucine (below)

Gel filtration experiments showed that the variants are monomers, meaning four-helix bundles, in solution. The chromatographs are shown below in Figure 4.9.

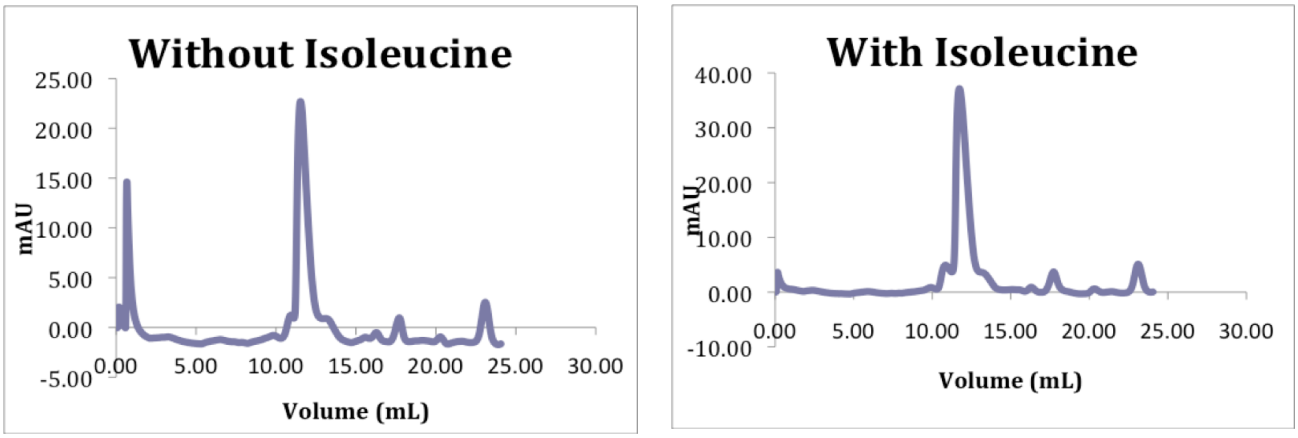


Figure 4.9. Gel filtration results of single-chain variants. Comparison to a standard curve indicates that they are four-helix bundles, or monomers, in solution.

Finally, crystals of each variant were obtained, and a data set was collected at Eli Lilly beamline for the variant without isoleucine. The structure is currently being solved by molecular replacement. Pictures of the crystallized single chain variants are shown below in Figure 4.10.

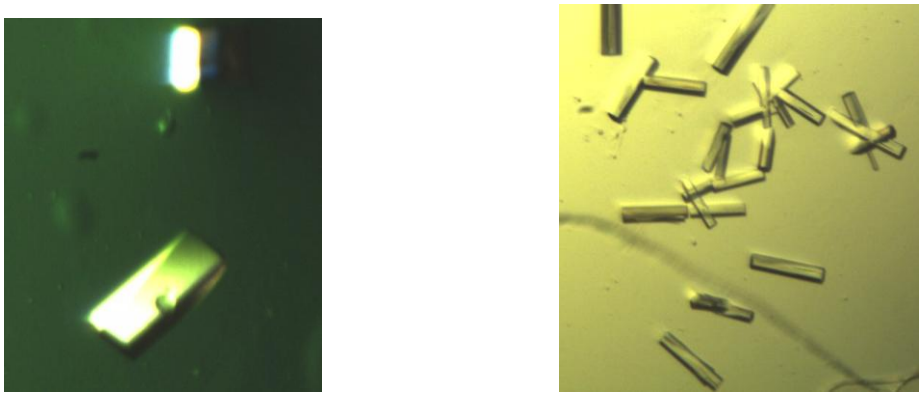


Figure 4.10. Crystals of single-chain Rop variants. The variant with isoleucine is shown to the left, and the variant without isoleucine is pictured to the right.

The Phe residue at position 14 in wild type Rop is of particular interest to researchers because it seems to have a huge effect on activity. All characterized Rop variants that are active in the screen have maintained the F14 position, and mutations to other residues do not give fluorescent phenotypes. When initially analyzing 1yo7, it was noticed that the F14 side chain sticks out into the solvent in a way that is not characteristic of the wild type (Figure 4.11). This oddity may be responsible for the lack of activity in 1yo7, and it will be interesting to see whether or not the modification of the loop residues in the engineered variants fixed this problem. The crystal structure will be able to give confirmation to this result, so solving the structures of these variants is a top priority. Future work includes creating single-chain libraries using one of these variants as a scaffold in order to analyze the effects of asymmetric mutations with a large number of sequences.

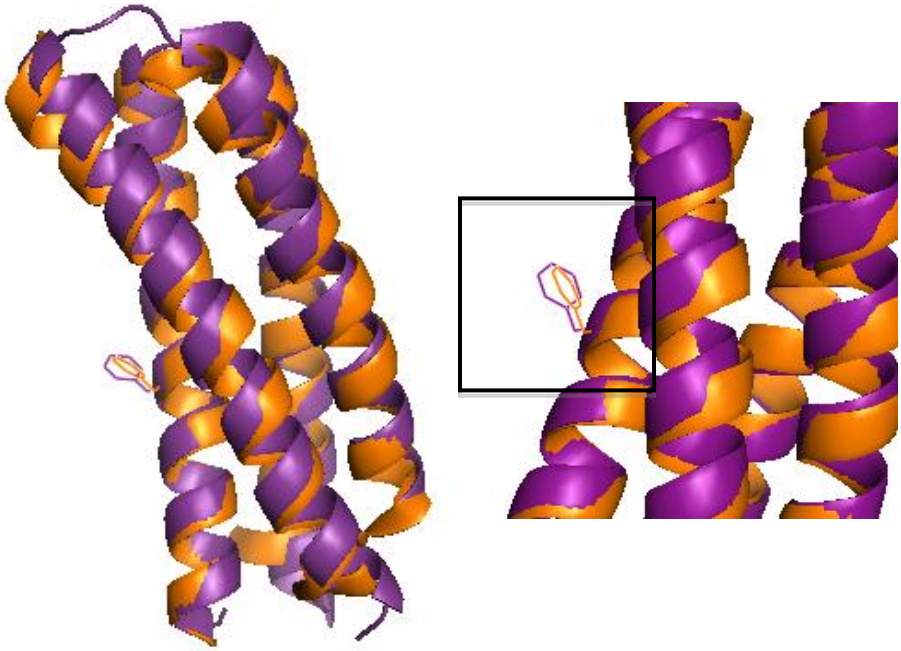


Figure 4.11. The orientation of the side chains of F14 in wild type Rop (orange) and 1yo7 (purple)

4.2 Biophysical characterization of Rop surface variants

The NNK_5 surface library cloned mutated positions 38, 39, 42, 45, and 46, which are all "b," "c," and "f" residues of the heptad repeat present in Rop bundles (Figure 4.12). These are the most solvent-exposed residues and are not present on the binding interface, which is why they were targeted for mutagenesis experiments. The original sequence, ESDDE, carried a highly negative net charge, and we were interested in what types of mutations could be tolerated at the surface. Figure 4.13 shows the positions of the 5 surface residues altered in this library.

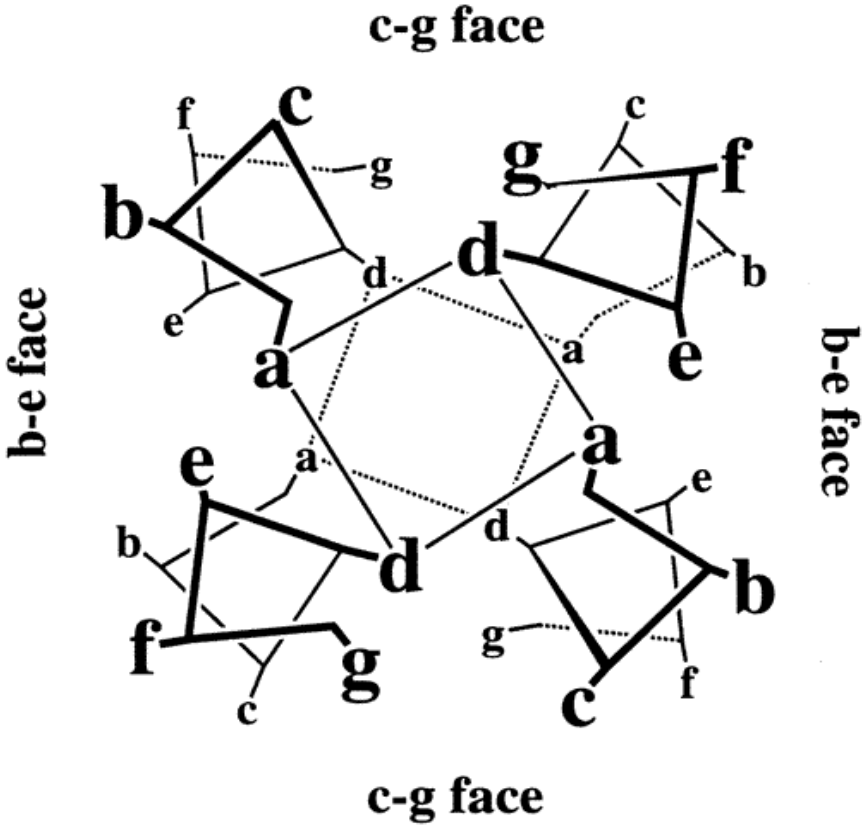


Figure 4.12. Heptad repeat diagram of two layers of Rop[43]

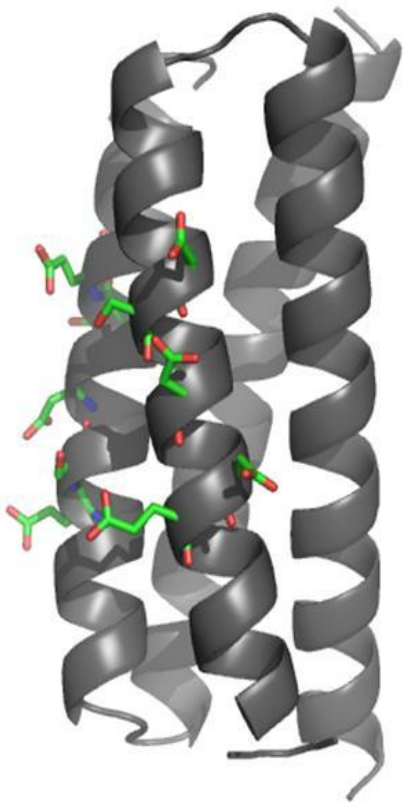
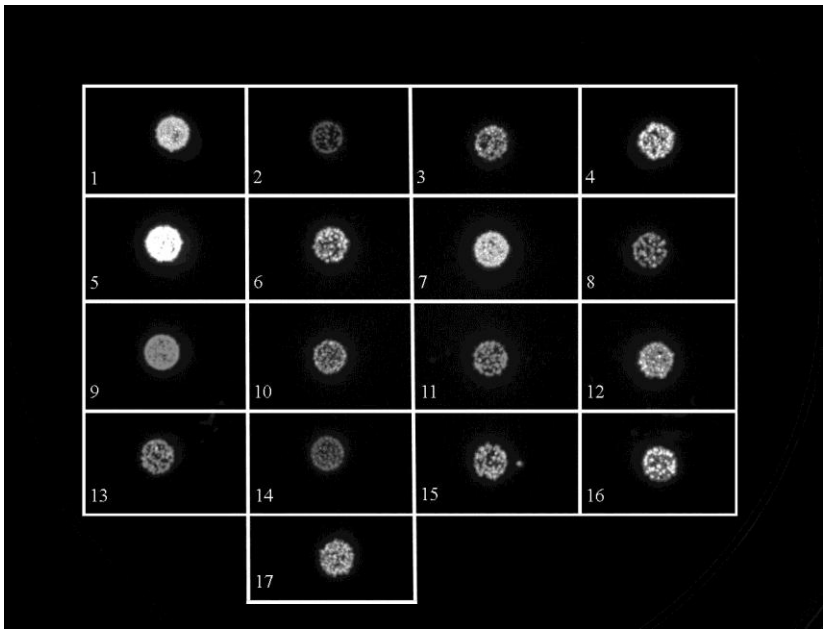


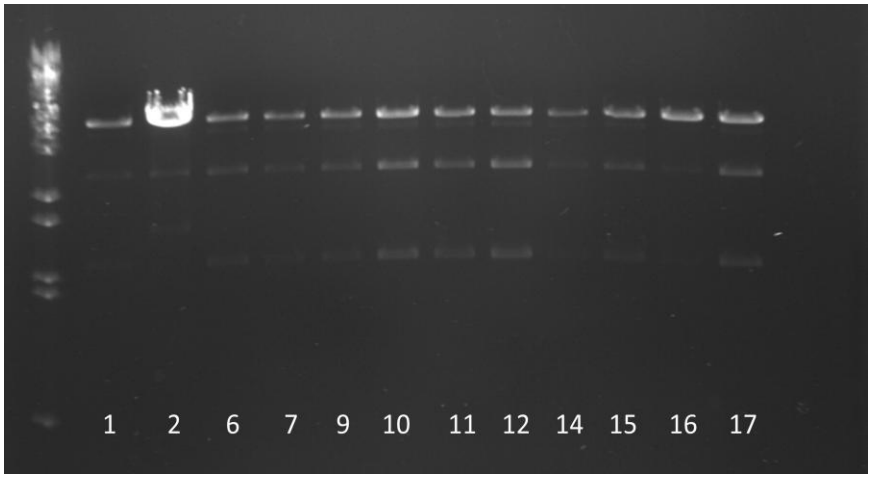
Figure 4.13. Rop diagram depicting the original side chains of the mutated surface residues

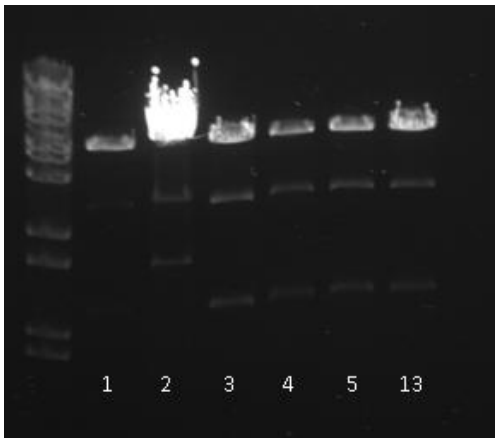
Of all the variants that passed the screen, the 16 Rop variants listed in the materials and methods section were chosen to undergo biophysical characterization. Figure 4.14 below shows the relative phenotypes of these variants on the *in vivo* activity screen, and their corresponding sequences are labeled. Their relative amounts of ColE1 plasmid were also compared by gel (Figure 4.15).



Variant	Sequence
1	ESDDE (wt)
2	Linker
3	VDWLL
4	LELSK
5	LMATA
6	CLCEL
7	YALDV
8	QHNIE
9	EETEC
10	DTAQQ
11	KTKKK
12	ELDAE
13	YSRHK
14	HTTEQ
15	ARFWE
16	EEDHM
17	DETHQ

Figure 4.14. Selected variants shown on *in vivo* activity screen





Variant	Sequence
1	ESDDE (wt)
2	Linker
3	VDWLL
4	LELSK
5	LMATA
6	CLCEL
7	YALDV
8	QHNIE
9	EETEC
10	DTAQQ
11	KTKKK
12	ELDAE
13	YSRHK
14	HTTEQ
15	ARFWE
16	EEDHM
17	DETHQ

Figure 4.15. Comparison of ColE1 plasmid levels by gel

The solubility of these variants was also analyzed through comparison of pellet and clear cell lysate. The results are shown below in Figure 4.16. It can be seen that the sequences LMATA, LELSK, VDWLL and YSRHK are barely found in the soluble fraction. Most of the protein is found in the pellet, so we decided to eliminate them from further characterization experiments based on our inability to work with them.

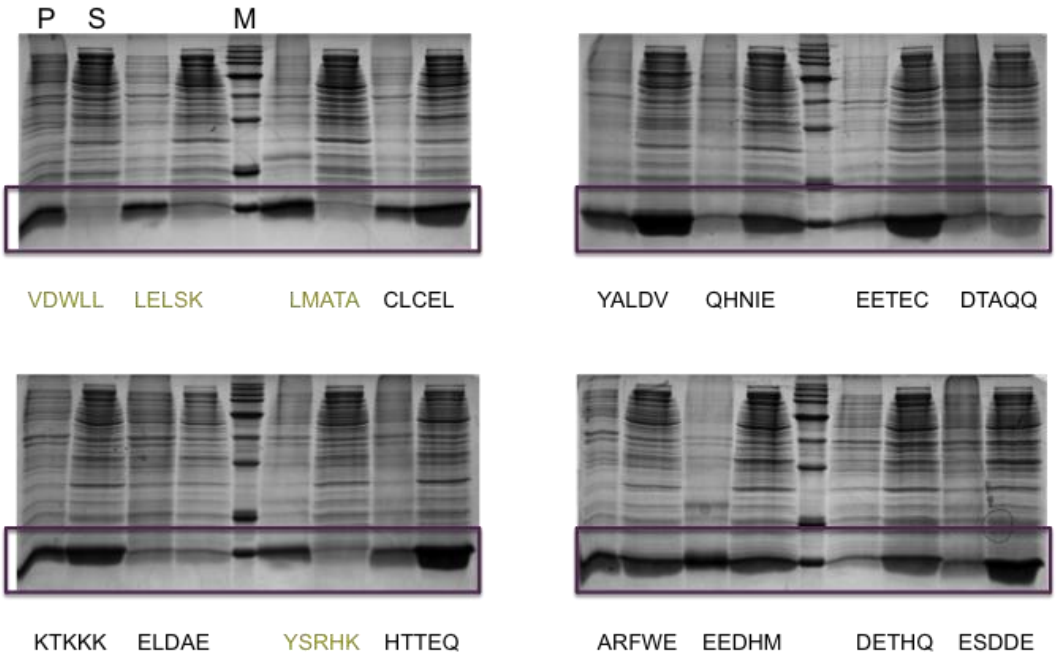


Figure 4.16. Expression tests showing the clear cell lysate and pellets of surface variants. Those in green were not found in the soluble fraction.

Of the remaining twelve variants, CD scans were taken and thermal denaturations were conducted. Peaks at 208 and 222 nm in (Figure 4.17) indicate the alpha-helical nature of the proteins, and the thermal melts are shown below in Figure 4.18. Scans and melts for the Cys-variants are not pictured, as problems with oxidative dimer formation have caused us to believe our former data might be invalid. The characterization of these variants is currently in the process of being completed.

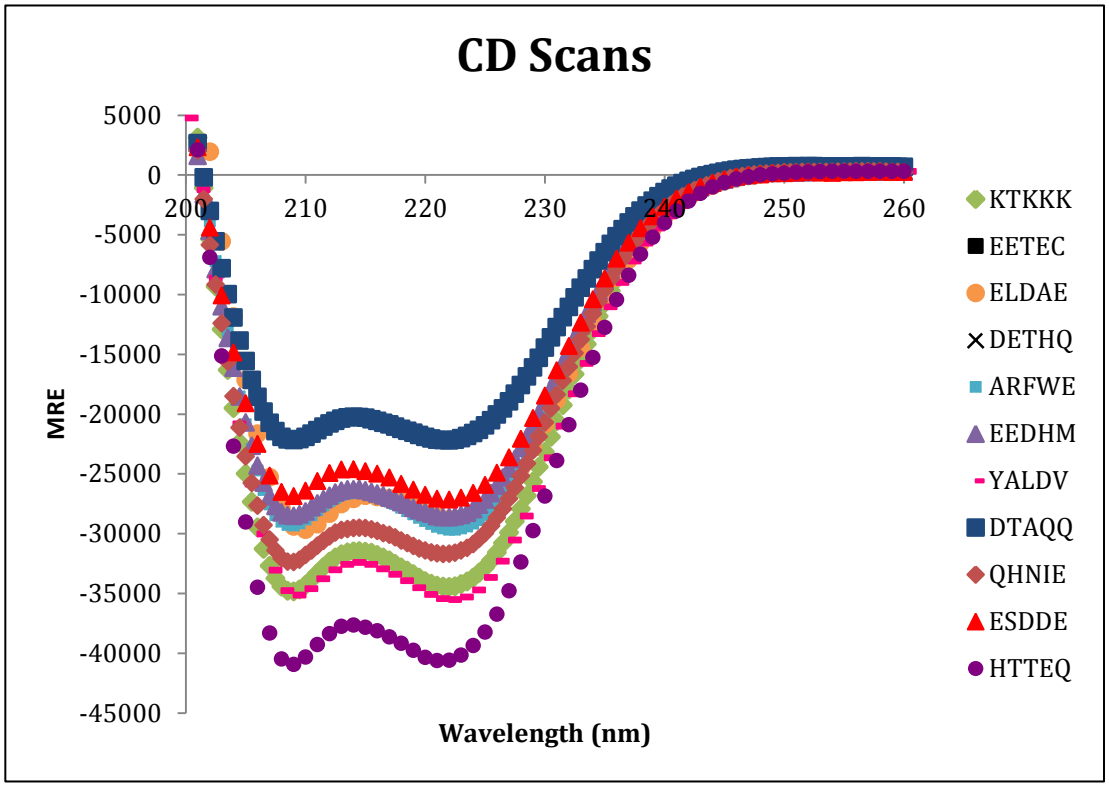


Figure 4.17. CD scans of Cys-free surface variants (wt shown in red triangles)

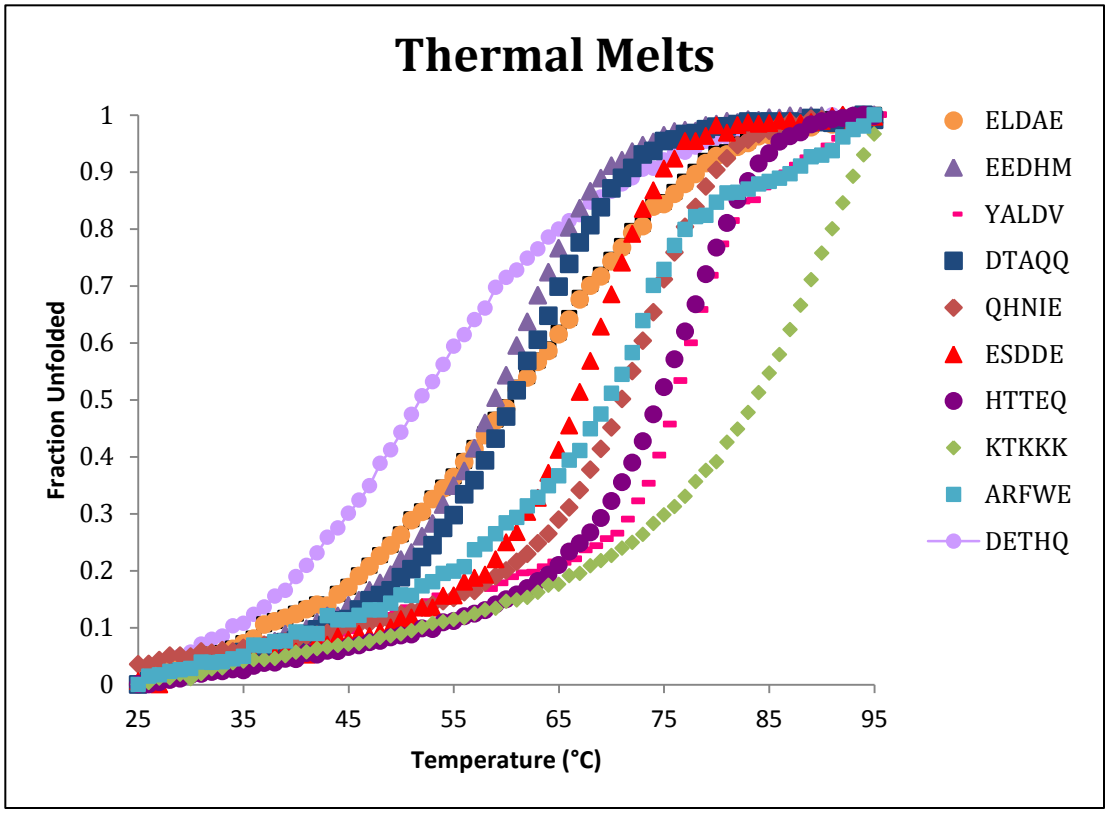


Figure 4.18. Thermal melts of Cys-free surface variants

As seen above in Figure 4.18, five variants were thermostabilized compared to the wild type sequence. It is especially interesting that variants such as ARFWE and YALDV show improved thermodynamic stability, considering the large amount of hydrophobic residues adopted at the surface. Another curious result is that of KTKKK, which never seems to completely unfold even at 95 °C. This highly charged sequence may be favorable at the solvent-exposed region, and current experiments such as urea denaturation are being conducted to determine a T_m for this variant.

Variant	Tm (°C)	MW (kD)
ESDDE	68.5	21.10
KTKKK	-	17.62
ELDAE	66.0	21.00
YALDV	76.2*	21.62
HTTEQ	78.4	23.26
QHNIE	73.6	22.44
EEDHM	60.1*	19.28
DTAQQ	62.9	24.16
DETHQ	50.6*	20.70
EETEC	-	20.63
CLCEL	-	-
ARFWE	71.8	-

Table 4.1. A summary of results from surface library characterization. T_m values marked with * indicate irreversible unfolding and are technically $T_{1/2}$ values

The importance of surface mutations to proteins is suggested by the results of the characterization of this library. Sequences with multiple additions of hydrophobic residues were able to pass the activity screen, which would not be predicted based on our general knowledge of electrostatic interactions. In fact, the variants YALDV and LMATA showed some of the most fluorescent phenotypes on the screen (Figure 25), indicating high levels of Rop activity. It was surprising enough that such mutations were tolerated at the surface region, but even more intriguing is that some actually seemed to improve the activity and thermostability of the protein. Future work on this library includes the characterization of Cys-variants through CD and thermal melts, as well as optimizing crystallography conditions for potential structure determination.

References

1. Anfinsen, C. B., The formation and stabilization of protein structure. *Biochem. J.* **1972,** 128, (4), 737-749.
2. Richards, F. M., Protein stability: still an unsolved problem. *CMLS* **1997,** 53, 790-802.
3. Magliery, T.J., Regan, L., Combinatorial approaches to protein stability and structure. *Eur. J. Biochem.* **2004**, 271, 1595-1608.
4. Dill, K.A., Chan, H. S., From Levinthal to pathways to funnels. *Nat. Struct. Biol.* **1997,** 4, (1), 10-19.
5. Dill, K. A. et al. The protein folding problem: When will it be solved? *Curr. Opin. Struct. Biol.* 2007, 17, (3), 342-346.
6. Yue, K., Dill, K.A. Inverse protein folding problem: Designing polymer sequences. *PNAS* **1992,** 89, 4163-4167.
7. Dahiyat, B. I., Mayo, S. L., De novo protein design: fully automated sequence selection. *Science* **1997,** 278, 82.
8. Lehmann, A., Saven, J. Computational design of four-helix bundle proteins that bind nonbiological cofactors. *Biotechnol. Prog.* **2008**, 24, 74-79.
9. DeGrado et al. Computational design and characterization of a monomeric helical dinuclear metalloprotein. *J. Mol. Bio.* **2003,** 334, (5), 1101-1115.

10. Hecht et al. Stably folded de novo proteins from a designed combinatorial library. *Protein Science* **2003,** 12, 92-102.
11. Lin, H., Cornish, V. W., Screening methods for large-scale analysis of protein function. *Angew. Chem. Int. Ed.* **2002,** 41, 4402-4425.
12. Cesareni, G., et al., Control of ColE1 DNA replication: the Rop gene product negatively affects transcription from the replication primer promoter. *Proc. Nat'l. Acad. Sci. U.S.A.* **1982,** 79, (20), 6313-6317.
13. Castagnoli, L., et al., Genetic and structural analysis of the ColE1 Rop (Rom) protein. *Embo. J.* **1989,** 8, (2), 621-629.
14. Banner, D. W., et al., Crystallization of the ColE1 Rop protein. *J. Mol. Biol.* **1983,** 170, (4), 1059-1060.
15. Banner, D. W., et al., Structure of the ColE1 Rop protein at 1.7 .ANG. resolution. *J. Mol. Biol.* **1987,** 196, (3), 657-75.
16. Hari et al. Cysteine-free Rop: A four-helix bundle core mutant has wild-type stability and structure but dramatically different unfolding kinetics. *Protein Science* **2010,** 19, 670-679.
17. Magliery, T. J.; Regan, L., A cell-based screen for function of the four-helix bundle protein Rop: a new tool for combinatorial experiments in biophysics. *Protein. Eng. Des. Sel.* **2004,** 17, (1), 77-83.
18. Lim, K. H., Huang, H., Pralle, A. Park, S., Engineered Streptavidin monomer and dimer with improved stability and function. *Biochemistry* **2011,** 50, (40), 8682-8691.

19. Mehl et al., Probing dimer interface stabilization within a four-helix bundle of the GrpE protein from Escheria coli via internal deletion mutants: Conversion of a dimer to a monomer. *Int. Journal of Biol. Macromolecules* **2011,** 48, 627-633.
20. Sander, C. Design of protein structures: Helix bundles and beyond. *Trends in Biotechnology* **1994,** 12, 163-167.
21. Predki, P. F., Regan, L. Redesigning the topology of a four-helix bundle protein: Monomeric Rop. *Biochemistry* **1995,** 34, 9834-9839.
22. Kresse et al. Four-helix bundle topology re-engineered: Monomeric Rop protein variants with different loop arrangements. *PEDS* **2001,** 14, (11), 897-901.
23. Westerlund et al. Making a single-chain four-helix bundle for redox chemistry studies. *Protein Engineering, Design and Selection* **2008,** 21, (11), 645-652.
24. Hecht et al. De novo proteins from designed combinatorial libraries. *Protein Sci.* **2004,** 13, 1711–1723.
25. Go et al. Structure and dynamics of de novo proteins from a designed superfamily of four-helix bundles. *Protein Science* **2008,** 17, 821-832.
26. Jung, S., Pluckthun, A. Improving in vivo folding and stability of a single-chain Fv antibody fragment by loop grafting. *Protein Engineering* **1997,** 10, (8), 959-966.
27. Helms, L. R., Wetzel, R. Destabilizing loop swaps in the CDRs of an immunoglobulin V_L domain. *Protein Science* **1995,** 4, 2073-2081.

28. Magalhães et al. Evolved Streptavidin mutants reveal key role of loop residue in high-affinity binding. *Protein Science* **2011,** 20, (7), 1145-1154.

29. Predki et al. Amino-acid substitutions in a surface turn modulate protein stability. *Nature Structural Biology* **1996,** 3, (54), 54-60.

30. Nagi, A., Regan, L. An inverse correlation between loop length and stability in a four-helix bundle protein. *Folding and Design* **1997,** 2, 67-75.

31. Gibney et al. Using loop length variants to dissect the folding pathway of a four-helix bundle protein. *J. Mol. Biol.* **1999**, 286, 257-265.

32. Kauzmann, W. J., Factors in the interpretation of protein denaturation. *Adv. Prot. Chem.*, **1959,** 14, 1-53.

33. Dill, K., Dominant forces in protein folding. *Biochemistry* **1990,** 29, (31), 7133-7155.

34. Schueler-Furman, O., Wang, C., Bradley, P., Misura, K., Baker, D., Progress in modeling of protein structures and interactions. *Science* **2005,** 310, (5748), 638-642.

35. Luo, P., Baldwin, R. L., Interaction between water and polar groups of the helix backbone: An important determinant of helix propensities. *Proc. Nat'l. Acad. Sci. USA* **1999,** 96, 4930-4935.

36. Anderson, D. E., Becktel, W. J., Dahlquist, F.W., *Biochemistry* **1990,** 29, 2403-2408.

37. Schreiber, G., Buckle, A. M., Fersht, A. R. *Structure* **1994,** 2, 945-951.

38. Koide et al. Stabilization of a fibronectin type III domain by the removal of unfavorable electrostatic interactions on the protein surface. *Biochemistry* **2001,** 40, 10326-10333.
39. Spector et al. Rational modification of protein stability by the mutation of charged surface residues. *Biochemistry* **2000,** 39, 872-879.
40. Ibarra-Molero et al. Thermal vs. guanidine-induced unfolding of ubiquitin. An analysis in terms of the contributions from charge-charge interactions to protein stability. *Biochemistry* **1999,** 28, 8138-8149.
41. Makhatadze et al. Contribution of surface salt bridges to protein stability: Guidelines for protein engineering. *J. Mol. Biol.* **2003,** 327, 1135-1148.
42. Perl et al. Two exposed amino acid residues confer thermostability on a cold shock protein. *Nature Structural Biology* **2000** 7, (5), 380-383.
43. Betz, S. F., Liebman, P. A., DeGrado, W.F. *De novo* design of native proteins: characterization of proteins intended to fold into antiparallel, Rop-like, four-helix bundles. *Biochemistry* **1997,** 26, 2450-2458.